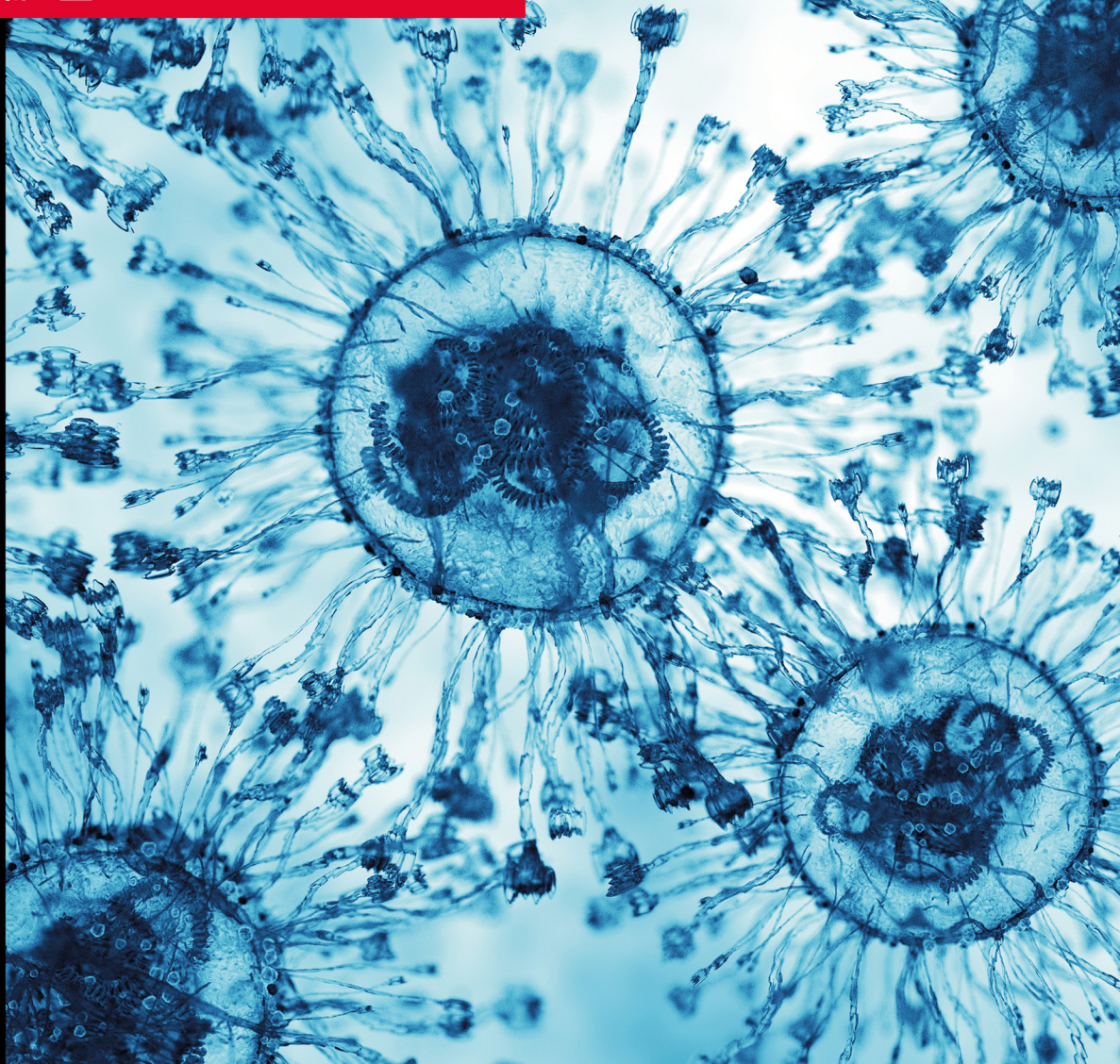




ROUTLEDGE
HANDBOOKS



The Routledge Handbook of Values and Science

Edited by Kevin C. Elliott and Ted Richards



THE ROUTLEDGE HANDBOOK OF VALUES AND SCIENCE

This is the first-ever handbook to cover the vibrant philosophical literature on values and science.

Its 45 chapters—appearing in print here for the first time—were written by a distinguished, international group of contributors and have been organized into six parts that explore the many ways in which ethics, values, and social considerations relate to the practices and content of the sciences:

1. Theoretical Background on Values and Science
2. Managing Values in Science
3. Values, Science, and Democracy
4. Values, Science, Institutions, and Organizations
5. Values in Scientific Activities
6. Values in Specific Sciences

Published at a time of increasing concerns about misinformation, polarization, and lack of reproducibility in science, this volume is essential reading for anyone interested in understanding how science can be made more ethically and socially responsible.

The Routledge Handbook of Values and Science is designed to be an accessible resource not only for philosophers but also for scientists, policy makers, science communicators, and scholars from other science-studies fields.

Kevin C. Elliott is Red Cedar Distinguished Professor with joint appointments in Lyman Briggs College, the Department of Fisheries and Wildlife, and the Department of Philosophy at Michigan State University. His books include *Values in Science* (2022), *A Tapestry of Values: An Introduction to Values in Science* (2017), and *Is a Little Pollution Good for You? Incorporating Societal Values in Environmental Research* (2011).

Ted Richards is Academic Specialist in the Department of Philosophy at Michigan State University. He is the editor of *The Rightful Place of Science: Science, Values and Democracy* (2021), *Exploring Inductive Risk* (2017) with Kevin C. Elliott, and *Soccer and Philosophy: Beautiful Thoughts on the Beautiful Game* (2010).

ROUTLEDGE HANDBOOKS IN PHILOSOPHY

Routledge Handbooks in Philosophy are state-of-the-art surveys of emerging, newly refreshed, and important fields in philosophy, providing accessible yet thorough assessments of key problems, themes, thinkers, and recent developments in research.

All chapters for each volume are specially commissioned, and written by leading scholars in the field. Carefully edited and organized, *Routledge Handbooks in Philosophy* provide indispensable reference tools for students and researchers seeking a comprehensive overview of new and exciting topics in philosophy. They are also valuable teaching resources as accompaniments to textbooks, anthologies, and research-orientated publications.

Also Available:

THE ROUTLEDGE HANDBOOK OF MINDSHAPING

Edited by Tadeusz Wiesław Zawidzki and Rémi Tison

THE ROUTLEDGE HANDBOOK OF PHILOSOPHY OF HUMAN RIGHTS

Edited by Jesse Tomalty and Kerri Woods

THE ROUTLEDGE HANDBOOK OF THE HISTORY OF PHILOSOPHY OF SCIENCE AFTER KANT

Edited by Flavia Padovani and Ádám Tamás Tuboly

THE ROUTLEDGE HANDBOOK OF ARGUMENTATION THEORY

Edited by Scott Aikin, John Casey, and Katharina Stevens

THE ROUTLEDGE HANDBOOK OF VALUES AND SCIENCE

Edited by Kevin C. Elliott and Ted Richards

For more information about this series, please visit: <https://www.routledge.com/Routledge-Handbooks-in-Philosophy/book-series/RHP>

THE ROUTLEDGE HANDBOOK OF VALUES AND SCIENCE

Edited by Kevin C. Elliott and Ted Richards



ROUTLEDGE

Routledge
Taylor & Francis Group

NEW YORK AND LONDON

Designed cover image: Getty Images

First published 2026

by Routledge

605 Third Avenue, New York, NY 10158

and by Routledge

4 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2026 selection and editorial matter, Kevin C. Elliott and Ted Richards;
individual chapters, the contributors

The right of Kevin C. Elliott and Ted Richards to be identified as the authors
of the editorial material, and of the authors for their individual chapters,
has been asserted in accordance with sections 77 and 78 of the Copyright,
Designs and Patents Act 1988.

The Open Access version of this book, available at www.taylorfrancis.com,
has been made available under a Creative Commons Attribution-Non
Commercial-No Derivatives (CC BY-NC-ND) 4.0 International license.

Any third party material in this book is not included in the OA Creative
Commons license, unless indicated otherwise in a credit line to the material.

Please direct any permissions enquiries to the original rights holder.

For Product Safety Concerns and Information please contact our EU
representative GPSR@taylorandfrancis.com. Taylor & Francis Verlag GmbH,
Kaufingerstraße 24, 80331 München, Germany.

Trademark notice: Product or corporate names may be trademarks or registered
trademarks, and are used only for identification and explanation without intent
to infringe.

ISBN: 978-1-032-74413-1 (hbk)

ISBN: 978-1-032-74415-5 (pbk)

ISBN: 978-1-003-46910-0 (ebk)

DOI: [10.4324/9781003469100](https://doi.org/10.4324/9781003469100)

Typeset in Times New Roman
by KnowledgeWorks Global Ltd.

To our loving partners, whom we value above all, and who have suffered through this second book project. We'd like to say that this is the last, but....

—KCE & TR



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

CONTENTS

<i>List of Contributors</i>	<i>xii</i>
<i>Acknowledgments</i>	<i>xviii</i>
1 An Introduction to Values and Science	1
<i>Kevin C. Elliott and Ted Richards</i>	
PART I	
Theoretical Background on Values and Science	21
2 Feminist Challenges to the Value-Free Ideal of Science	23
<i>Kristen K. Intemann</i>	
3 Standpoint Theory	34
<i>Kristina Rolin</i>	
4 Values and Objectivity	42
<i>Inkeri Koskinen</i>	
5 Arguments Against the Value-Free Ideal	52
<i>Kevin C. Elliott</i>	
6 Recent Arguments for the Ideal of Value-Free Science	63
<i>Matthew J. Brown</i>	
7 What Does It Mean to Say that Science is Value-Laden?	74
<i>Zina B. Ward</i>	

Contents

8	Aesthetic Values in Scientific Practice: Contemporary Debates <i>Milena Ivanova</i>	84
9	Representational Risk Revisited <i>Stephanie Harvard and Eric Winsberg</i>	95
PART II		
	Managing Values in Science	107
10	Reflecting on Responses to the New Demarcation Problem <i>Bennett Holman and T.Y. Branch</i>	109
11	An Intermediate Approach to Value Management <i>Wendy S. Parker</i>	120
12	Democracy, Consensus, and the Value-Free Ideal <i>Tarun Menon and Jacob Stegenga</i>	131
13	Critical Contextual Empiricism <i>Kirstin Borgerson</i>	141
14	Values in Global Science and the Relevance of Geographic Diversity <i>Juliana Gutiérrez Valderrama</i>	151
15	The Limits of Diversity in Science: The Case of Human Microbiome Research <i>Abigail Nieves Delgado, Jan Baedke, and Ann-Christin Fischer</i>	162
16	Transparency in Science <i>Stephen John</i>	175
PART III		
	Values, Science, and Democracy	187
17	Scientific Knowledge as a Public Resource: Arguments and Challenges for a Democratic Approach to Values in Science <i>S. Andrew Schroeder</i>	189
18	Measuring the Public's Values: Are the Data Adequate for Purpose? <i>Karen Kovaka</i>	199
19	FDA Evidentiary Standards and the Need to Attend to Stakeholders' Values <i>Inmaculada de Melo-Martín</i>	209

Contents

20	Science and Democracy	220
	<i>Hannah Hilligardt and Torsten Wilholt</i>	
21	Science and Religion in a Democratic Society	231
	<i>Yiftach Fehige</i>	
22	Values in Science in Nondemocratic Contexts	241
	<i>Karoliina Pulkkinen</i>	
23	Science and Social Justice	252
	<i>Heather Douglas</i>	
24	Algorithmic Abolitionism and the Racial Algorithm	263
	<i>Alexander Williams Tolbert</i>	
PART IV		
Values, Science, Institutions, and Organizations		279
25	The Values of Science Funding Institutions	281
	<i>Jamie Shaw</i>	
26	Values and Industry-Funded Research	294
	<i>Pedro Bravo</i>	
27	Epistemic Intimidation and Illegitimate Value Influences in Science	304
	<i>Manuela Fernández Pinto and Anna Leuschner</i>	
28	Networks and Values	315
	<i>Rebecca Korf and Cailin O'Connor</i>	
29	Valuing Partnerships in Agricultural Genomics Research: Using More Empirically Informed Philosophical Approaches to Advance the Science and Values Debates	326
	<i>Rachel A. Ankeny and Emily A. Buddle</i>	
30	Values and Assessment Reports on Climate Change	337
	<i>Ahmad Elabbar</i>	
31	Indigenous Peoples' Leadership in Environmental Science Assessments	348
	<i>Kyle Whyte and Pasang Yangjee Sherpa</i>	

Contents

32	Controversies in Vaccine Policy <i>Maya J. Goldenberg</i>	363
33	Values and Dual Use Biomedical Research <i>David B. Resnik</i>	374
PART V		
	Values in Scientific Activities	387
34	Values and Measurement <i>Cristian Larroulet Philippi</i>	389
35	Science, Values, and Race Correction <i>Zinhle Mncube</i>	401
36	Value Overlaps in Transdisciplinary and Intercultural Collaboration <i>Charbel N. El-Hani and David Ludwig</i>	413
37	Tragedy or Transition?: How Science and Values Matter for Climate Change Frames <i>Daniel Steel</i>	425
38	Values in Science Communication Models <i>T.Y. Branch</i>	435
39	Values in Research Ethics Education <i>Kristina Rolin</i>	444
PART VI		
	Values in Specific Sciences	455
40	“Every Cell Has a Sex”: Sex Essentialism in Biomedical Research <i>Robyn Bluhm</i>	457
41	Values in Human Genomics <i>Yasmin Haddad and Celso Neto</i>	467
42	Philosophical Approaches to Values in Climate Science <i>Greg Lusk</i>	478

Contents

43	The Power of Value-Laden Framing: Examples from Wildlife Conservation Research	489
	<i>Jacalyn M. Beck, Claire F. Hoffmann, Kelly Kapsar, and Robert A. Montgomery</i>	
44	Values in Economics	501
	<i>Elizabeth Anderson</i>	
45	Artificial Intelligence: Values, Governance, and Policy	511
	<i>Justin B. Biddle</i>	
	<i>Index</i>	523

CONTRIBUTORS

Elizabeth Anderson is John Dewey Distinguished University Professor of Philosophy and Women's and Gender Studies at University of Michigan, Ann Arbor (USA). She is the author of *Value in Ethics and Economics* (Harvard University Press, 1993), *Private Government: How Employers Rule Our Lives (and Why We Don't Talk About It)* (Princeton University Press, 2017), and *Hijacked: How Neoliberalism Turned the Work Ethic against Workers and How Workers Can Take It Back* (Cambridge University Press, 2023).

Rachel A. Ankeny is Professor and Chair of Philosophy at Wageningen University (The Netherlands) and Affiliated Professor at University of Adelaide (Australia) and does research in the philosophy of science in practice, philosophy of the biological and biomedical sciences, and ethics of food and agriculture.

Jan Baedke is Professor at the Department of Philosophy I, Ruhr University Bochum (Germany). His research is characterized by an integrated approach to the history and philosophy of the life sciences, with a focus on evolutionary biology, microbiology, and the concept of race. He is the author of *The Organism* (Cambridge University Press, 2025) and *Above the Gene, Beyond Biology: Toward a Philosophy of Epigenetics* (University of Pittsburgh Press, 2018) and the PI of the German Research Foundation-funded research group "ROTO" (The Return of the Organism in the Bio-sciences: Theoretical, Historical, and Social Dimensions).

Jacalyn M. Beck is Division Director with the American Conservation Experience's Emerging Professionals in Conservation program (USA), where she manages youth internships in partnership with land management agencies across the United States. Her scholarly pursuits aim to increase the integration of philosophy and social science techniques into conservation research, education, and practice.

Justin B. Biddle is Associate Professor in the Jimmy and Rosalynn Carter School of Public Policy at the Georgia Institute of Technology, Atlanta, Georgia (USA).

Robyn Bluhm is Professor in the Department of Philosophy and Lyman Briggs College at Michigan State University (USA).

Contributors

Kirstin Borgerson is Associate Professor of Philosophy (cross-appointed to Bioethics and Gender & Women's Studies) at Dalhousie University (Canada).

T.Y. Branch is a postdoctoral researcher with the SOCRATES Centre for Advanced Studies based at the Institute for Philosophy at Leibniz University Hannover (Germany).

Pedro Bravo is Assistant Professor of Epistemology and Philosophy of Science at the Federal University of ABC (Brazil).

Matthew J. Brown is Jo Ann and Donald N. Boydston Chair of American Philosophy, Director of the Center for Dewey Studies, and Professor of Philosophy at Southern Illinois University Carbondale (USA). He is the author of *Science and Moral Imagination: A New Ideal for Values in Science* (University of Pittsburgh Press, 2020), the editor of *HOPPOS: The Journal of the International Society for the History of Philosophy of Science*, and the founder of the *Values in Medicine, Science, and Technology Conference*.

Emily A. Buddle is Senior Research Fellow in the School of Agriculture, Food and Wine at the University of Adelaide (Australia). Her research is interdisciplinary, centered in the broad areas of social and ethical issues in food, agriculture, and science.

Inmaculada de Melo-Martín is Professor of Medical Ethics in the Division of Medical Ethics at Weill Cornell Medical College (USA). Her latest books are *Rethinking Reprogenetics* (Oxford University Press, 2017) and, with K. Intemann, *The Fight Against Doubt* (Oxford University Press, 2018).

Heather Douglas is Professor in the Department of Philosophy at Michigan State University (USA). Her books include *Science, Policy and the Value-free Ideal* (University of Pittsburgh Press, 2009), *Environmental Ethics from the Roots Up: An Introductory Anthology* (Cognella Press, 2016), and *Science, Values and Democracy: The 2016 Descartes Lectures* (Consortium for Science, Policy & Outcomes, 2021).

Ahmad Elabbar is Assistant Professor in Philosophy of Science and Technology at the University of Cambridge (UK). His research focuses on the ethics and politics of climate science.

Charbel N. El-Hani is Full Professor of History and Philosophy of Science at the Institute of Biology, Federal University of Bahia (Brazil). He published *Transformative Transdisciplinarity: An Introduction to Community-Based Philosophy* (OUP, 2025) with David Ludwig, and *Conceptual Profiles: A Theory of Teaching and Learning Scientific Concepts* (Springer, 2014) with Eduardo F. Mortimer.

Kevin C. Elliott is Red Cedar Distinguished Professor at Michigan State University (USA). His research focuses on the roles of ethical and social values in the environmental health sciences. His books include *A Tapestry of Values: An Introduction to Values in Science* (Oxford University Press, 2017) and *Values in Science* (Cambridge University Press, 2022).

Yiftach Fehige is Full Professor for the history and philosophy of science and religion at the University of Toronto's IHPST (Canada). His most recent book is *Thought Experiments, Science, and Theology* (Brill, 2023), which was published in the series *Philosophical Studies in Science and Religion*.

Manuela Fernández Pinto is Associate Professor with tenure and Chair of the Department of Philosophy at Universidad de los Andes (Colombia). She is also Research Associate in the African Centre for Epistemology and Philosophy of Science at the University of Johannesburg. Together with Uskali Maki and Adrian Walsh, she is the editor of *Scientific Imperialism: Exploring the Boundaries of Interdisciplinarity* (Routledge, 2018).

Ann-Christin Fischer is a PhD candidate in Philosophy of Science at the Department of Philosophy I, Ruhr-University Bochum (Germany). Her research focuses on quality control in science and the role of values in science.

Maya J. Goldenberg is Professor of Philosophy at the University of Guelph (Canada). She is the author of *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science* (University of Pittsburgh Press, 2021).

Juliana Gutiérrez Valderrama is an independent researcher. Her interests include political and feminist philosophy of science, the science and values debate, and social and feminist epistemology. Her research focuses on the geopolitical dimensions of the production of scientific knowledge and the related epistemic and social consequences.

Yasmin Haddad is Postdoctoral Fellow funded by the Fonds de Recherche du Québec—Société et Culture at the Department of Philosophy at the Université du Québec à Montréal (Canada). Her work focuses on philosophy of science, philosophy of biology, and data ethics. Her project focuses on the concept of “ancestry” and its use in human population genetics.

Stephanie Harvard is Senior Scientist at Legacy for Airway Health in the Division of Respiratory Medicine at the University of British Columbia (Canada). She is the co-author of *Scientific Models and Decision-Making* (Cambridge University Press, 2024).

Hannah Hilligardt is a postdoc at the University of Bern (Switzerland). She works on questions related to science and democracy, values in science, and philosophy of climate science.

Claire F. Hoffmann is Head of Partnerships for the Yale Center for Biodiversity and Global Change at Yale University (USA) where her work centers on the research-implementation gap in conservation science. She helps to connect global-scale biodiversity data and the organizations, governments, and practitioners who can use that knowledge to inform more effective conservation decisions.

Bennett Holman is Associate Professor at Underwood International College, Yonsei University (South Korea).

Kristen K. Intemann is Professor of Philosophy and Director for the Center for Science, Technology, Ethics, and Society at Montana State University (USA). She is the co-author of *The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public* (Oxford University Press, 2018) with Inmaculada de Melo-Martín and *Feminist Epistemology and Philosophy of Science: An Introduction* (Routledge, 2024) with Sharon Crasnow.

Milena Ivanova is a philosopher of science working at the University of Cambridge (UK). She is Course Director in Philosophy at the Institute for Professional and Continuing Education and Senior

Teaching Associate at the Institute for Humanity and Technology. She is also Bye-Fellow at Newnham College. Ivanova is the author of *Duhem and Holism* (Cambridge University Press, 2021) and coeditor of *The Aesthetics of Science* (2020) and *The Aesthetics of Scientific Experiment* (2023), both published by Routledge.

Stephen John is Hatton Trust Professor of Philosophy of Public Health in the Department of History and Philosophy of Science, University of Cambridge (UK). He has published widely on public health ethics and philosophy of science, including his recent *Objectivity in Science* (Cambridge University Press, 2021).

Kelly Kapsar is Postdoctoral Research Associate in the Department of Integrative Biology at Michigan State University (USA) where she uses big data and supercomputing to study human-environment relationships from an interdisciplinary perspective. She works closely with practitioners in government and NGOs to identify management-relevant research questions and translate research findings into actionable outputs.

Rebecca Korf is Visiting Assistant Professor of Philosophy at Whitworth University (USA). Her research investigates the complex interactions between values and scientific research, especially given the social and collaborative nature of science.

Inkeri Koskinen is Academy of Finland Research Fellow in Practical Philosophy at the University of Helsinki (Finland).

Karen Kovaka is Assistant Professor of Philosophy at the University of California, San Diego (USA). Her research focuses on public participation in science and the epistemology of environmental science.

Cristian Larroulet Philippi is RW Seddon Fellow in Philosophy of Science at The University of Melbourne (Australia) and Junior Research Fellow at Gonville & Caius College, Cambridge (UK).

Anna Leuschner is Professor of Philosophy of Science and Deputy Director of the Interdisciplinary Center for Science and Technology Studies at the University of Wuppertal (Germany).

David Ludwig is Associate Professor at the Knowledge, Technology and Innovation (KTI) Group, Wageningen University & Research (The Netherlands). He published *Transformative Transdisciplinarity: An Introduction to Community-Based Philosophy* (OUP, 2025) with Charbel N. El-Hani, and *Global Epistemologies and Philosophies of Science* (Routledge, 2022) with Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo.

Greg Lusk is Assistant Professor of Environmental Philosophy at Durham University (UK) and is associate editor at *The British Journal for the Philosophy of Science*. He also co-directs Durham's Institute of Hazard Risk and Resilience as well as the Centre for Humanities Engaging Science and Society. Greg's work primarily focuses on two areas: Climate science and the interactions between science and democracy. His research is published in philosophy and science journals, including *Philosophy of Science*, *Bulletin of the American Meteorological Society*, and *Climatic Change*.

Tarun Menon is Associate Professor at Azim Premji University, Bengaluru (India).

Zinhle Mncube is Assistant Professor at the University of Massachusetts, Amherst (USA). She is a philosopher of science with an interest in race and personalizing medicine.

Robert A. Montgomery is Professor of Conservation Science, Director of the Centre for National Parks and Protected Areas, and Director of Research and Knowledge Exchange in the Institute of Science and Environment at the University of Cumbria (UK). He is the author of the edited volume *Apex Predators in the Anthropocene: How Humans Shape Large Carnivore Populations* (Oxford University Press, 2025).

Celso Neto is Associate Professor in Philosophy at the University of Exeter (UK), where he is part of the Centre for the Study of the Life Sciences (EGENIS). He works on the philosophy of biology and race, most recently as the principal investigator of a European Research Council project exploring the role of values and racial thinking in human genomics.

Abigail Nieves Delgado is Associate Professor at the Freudenthal Institute, Utrecht University (The Netherlands). Her research focuses on racialization practices in the history of science and in contemporary biomedical research (e.g., in microbiome research and epigenetics) and in biometric technologies (e.g., facial recognition). She is the principal investigator of the project “Microbiome research and race in the ‘Local South’” funded by the NWO.

Cailin O’Connor is a philosopher of science and Chancellor’s Professor of Logic and Philosophy of Science at the University of California, Irvine (USA). Cailin has written four books and published widely in both philosophy and the social sciences.

Wendy S. Parker is Professor of Philosophy at Virginia Tech (USA). Her research focuses on issues in general philosophy of science and philosophy of climate science. Most recently, she is the author of *Climate Science* (2025) in the Cambridge University Press *Elements in Philosophy of Science* series.

Karoliina Pulkkinen is a postdoctoral researcher in Aleksanteri Institute, University of Helsinki (Finland).

David B. Resnik is Bioethicist at the National Institute of Environmental Health Sciences and Senior Advisor for Research Integrity to the Intramural Research Program of the National Institutes of Health (USA). He has published ten books and over 300 articles on philosophical, ethical, and legal issues in science, technology, and medicine and is a AAAS Fellow.

Ted Richards is Academic Specialist in the Department of Philosophy at Michigan State University (USA). He is the editor of *Soccer and Philosophy* (Open Court, 2010), *Exploring Inductive Risk* (Oxford University Press, 2017, with Kevin C. Elliott), and *Science, Values and Democracy: The 2016 Descartes Lectures* (Consortium for Science, Policy & Outcomes, 2021).

Kristina Rolin is University Lecturer in Research Ethics at Tampere University (Finland). Her main areas of research are philosophy of science and social science, social epistemology, feminist epistemology, and research ethics.

S. Andrew Schroeder is Professor of Philosophy at Claremont McKenna College. His research focuses on questions at the intersection of the philosophy of science, political philosophy, and ethics.

Jamie Shaw is Postdoctoral Fellow at the Institute for Philosophy at Leibniz University Hannover (Germany). His primary interests include science funding policy, issues pertaining to scientific pluralism, and the history of philosophy of science (especially Paul Feyerabend). He is the coeditor of *Interpreting Feyerabend: Critical Essays* (Cambridge University Press, 2021) and *Values, Pluralism and Pragmatism: Themes from the Work of Matthew J. Brown* (Boston Studies in the History and Philosophy of Science, 2025).

Pasang Yangjee Sherpa is Assistant Professor of Lifeways in Indigenous Asia at the University of British Columbia, jointly appointed in the Institute for Critical Indigenous Studies and the Department of Asian Studies.

Daniel Steel is Associate Professor at the W. Maurice Young Centre for Applied Ethics in the School of Population and Public Health at the University of British Columbia, Vancouver (Canada). He is the author of *Across the Boundaries: Extrapolation in Biology and Social Science* (Oxford University Press, 2007) and *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy* (Cambridge University Press, 2015).

Jacob Stegenga is Professor at NTU (Singapore). His research focuses on philosophy of science and philosophy of medicine and he is the author of *Medical Nihilism* (Oxford University Press, 2018) and *Heart of Science* (The University of Chicago Press, 2025).

Alexander Williams Tolbert is Assistant Professor of Data and Decision Sciences at Emory University. He works in philosophy of science, applied ethics, and philosophy, politics, and economics (PPE). His research develops interdisciplinary approaches to artificial intelligence, causal inference, and ethics, with a focus on how algorithms shape decision-making in social systems.

Zina B. Ward is Assistant Professor of Philosophy at Florida State University (USA). She is the 2022 winner of the Popper Prize from the *British Journal for the Philosophy of Science*.

Kyle Whyte teaches at the University of Michigan, where he is George Willis Pack Professor in the School for Environment and Sustainability, University Diversity and Social Transformation Professor, Professor of Philosophy and Native American Studies, and Senior Fellow in the Michigan Society of Fellows. Kyle is Faculty Director of the Tishman Center for Social Justice and the Environment, the Environmental Justice + Humanities Hub, and the Environmental Justice Graduate Specialization.

Torsten Wilholt is Professor of Philosophy and History of the Natural Sciences at the Institute of Philosophy and the Leibniz Center for Science and Society at Leibniz Universität Hannover (Germany). He is co-directing the DFG-funded Centre for Advanced Studies “SOCRATES—Social Credibility and Trustworthiness of Expert Knowledge and Science-based Information.”

Eric Winsberg is British Academy Global Professor in the Department of History and Philosophy of Science at University of Cambridge and Professor of Philosophy at University of South Florida (USA). He is the author of *Science in the Age of Computer Simulation* (University of Chicago Press, 2010) and *Philosophy of Climate Science* (Cambridge University Press, 2018) and co-author of *Scientific Models and Decision-Making* (Cambridge University Press, 2024).

ACKNOWLEDGMENTS

We are grateful to many people and organizations for enabling us to accomplish a project of this magnitude. First, we thank Andrew Beck, our editor at Routledge, who developed the idea for this volume. Second, we are so thankful that all our authors accepted our invitation to participate in the handbook and produced such exciting contributions. We would also like to thank Heather Douglas for her generous ideas and input throughout the development of this book.

During his work on the volume, Kevin received support from a Research Award from the Alexander von Humboldt Foundation and from the SOCRATES Centre at Leibniz University, Hannover, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project 470816212/KFG43. He also benefited from a teaching leave made possible through the assistance of Kendra Spence Cheruvilil, Dean of Lyman Briggs College, and Dana Infante, Chair of the Department of Fisheries and Wildlife at Michigan State University.

Ted and Kevin are very grateful for funding provided by the SOCRATES Centre to make the book open access. Additional funding to cover the open access fees came from professional development funds provided to Ted by the Department of Philosophy at Michigan State University and a Humanities and Arts Research Program-Production grant awarded to Kevin by the Michigan State University Research Foundation.

1

AN INTRODUCTION TO VALUES AND SCIENCE

Kevin C. Elliott and Ted Richards

1.1 Introduction

At first glance, there appears to be little in common between science and values. Science is supposed to be about *facts*, which are true no matter what we hope or prefer to be the case. In contrast, values are supposed to be about what we *desire* or *prefer*. To put it simply, science appears to be about the way the world *is*, whereas values appear to be about what we *want* the world to be. Especially in today's era of misinformation, polarization, and "fake news," it might seem essential to keep these two realms as distinct from each other as possible. Only by keeping scientists focused on facts alone can we maintain the reliability and authority of science, so the thinking might go.

But a moment's reflection reveals that this distinction is too simple. We correctly applaud the researcher who studies cancer because they want to save lives. Notice that this desire to prevent human deaths reflects how they *want* the world to be, not how the world *is*. Thus, we can see that it is appropriate and even advisable for scientists to consider values when deciding what topics to study and how to study those topics. Upon further reflection, it also becomes clear that values have a crucial role to play when using science to inform decision making. For example, scientists can predict how likely it is for a nuclear power plant to malfunction, but values are important when deciding whether it is appropriate to accept that risk and build the plant. Even when scientists are considering how to communicate their findings, values are clearly relevant. Consider again the case of the nuclear power plant. When scientists predict the likelihood that it will malfunction, they have to draw on enormous amounts of data and make a wide variety of assumptions. Under such conditions, different experts are likely to come to different conclusions. Deciding how to present all those different conclusions in a responsible manner requires value-laden choices about how much to prioritize completeness as opposed to making information easy to understand.

The more one reflects on these situations, the more complex they get. Consider once again scientists trying to predict the likelihood that a nuclear power plant will malfunction. As they formulate their predictions, they face a wide variety of questions. Which data should they include in their analyses? Should they include safety data about all existing nuclear power plants, or only power plants that are sufficiently similar to the one under consideration? And how should they extrapolate from the existing data into the future? What if they suspect that conditions in the future might be different from those in the past? Depending on how the scientists answer these questions, they might overestimate or underestimate the likelihood of a malfunction. Which is a more tolerable error to make? Once

again, it seems like responsible scientists need to consider values in order to handle this decision. In other words, they need to reflect on how different errors could affect society, which is a decision about how they *want* the world to be, not about the way the world *is*.

One might argue that scientists could avoid considering values if they either refrained from making a prediction or if they just described a wide range of possible outcomes. This seems, however, unlikely to solve the problem. The scientists would still have to make value-laden decisions about when to withhold their judgment, what range of possibilities to describe, and/or how to describe the plausibility of different options. So we are left with difficult questions about how scientists can navigate all these roles for values in a responsible fashion that maintains scientific integrity.

Interest in these sorts of questions has exploded in recent decades among philosophers of science. This topic—frequently labeled “values and science”—has taken on even more significance because of contemporary concerns about misinformation (O’Connor and Weatherall 2019), corporate manipulation of science (Oreskes and Conway 2010), and problems with scientific reproducibility (Harris 2017). The COVID-19 pandemic brought even more attention to the relationship between values and science, as calls to “follow the science” often overlooked the important ways that values shaped scientific models and policy recommendations (Harvard et al. 2021; Macedo and Lee 2025). This volume provides an introduction to the contemporary literature on values and science and can serve as a resource both for those new to the debates and for those who have already been participating in them.

Contemporary work on values and science draws on several previous areas of research. First, the famous historian and philosopher of science Thomas Kuhn (1977) argued that reasonable scientists could legitimately disagree about how to weigh multiple different (e.g., explanatory power, predictive accuracy, and simplicity) when choosing among multiple scientific theories, such as Copernican vs Ptolemaic accounts of the solar system. His work launched subsequent discussions about whether it was possible to distinguish “epistemic” values, which provided evidence for a theory’s truth, from “non-epistemic values,” which failed to be truth-conducive (see, e.g., McMullin 1982). Second, feminist scholarship drew attention to the role that social, political, and ethical values could play in developing scientific theories and shaping scientists’ assumptions. As Kristen Intemann discusses in her chapter for this handbook, the work of these feminist scholars raised questions about whether value influences in science were inherently problematic or whether incorporating feminist values in science could sometimes generate better, more objective science (Intemann 2026; see also Kourany 2010). Helen Longino (1990, 2002) famously argued that scientists invariably rely on value-laden background assumptions when drawing conclusions, so it is better for them to subject these background assumptions and values to “transformative criticism” from diverse perspectives rather than pretending to be able to draw value-free conclusions. Third, Heather Douglas (2000, 2009) developed the “argument from inductive risk” (see Elliott 2026), which brought renewed attention to mid-twentieth-century philosophical discussions about the risks involved in making inductive inferences. Douglas argued that scientists have ethical responsibilities to consider the potential social consequences of drawing erroneous conclusions. On this basis, she concluded that the “value-free ideal” (VFI) for science, which had been prominent in previous philosophical thinking (e.g., McMullin 1982), should be rejected as a bad ideal.

Our aim in this handbook is to provide an accessible overview of the vibrant body of literature that has developed from these starting points. It would be unrealistic to try to represent this literature in a comprehensive fashion. Instead, we have sought to focus on promising new directions in which the field is headed while also providing readers with an introduction to the most important existing debates. This introductory chapter provides an orientation to key conceptual issues, clarifies how the various parts of the book fit together, and highlights important opportunities for further scholarship.

1.2 The nature of values and the value-free ideal

Values are generally thought to be concepts that reflect our ideals, preferences, or goals (see [Elliott and Korf \(2024\)](#) for further clarification). Values, as such, need not be fully conscious and can range over many kinds, whether epistemic, aesthetic, political, social, personal, or ethical. Matthew Brown has emphasized the distinction between values—which can be held without reflection—and value judgments—which are choices or conclusions that result from careful inquiry regarding which values we should hold, pursue, and prioritize ([Brown 2020](#)). Regardless of whether we are discussing values or value judgments, one key question concerns whether and how values should play a role in science.

During the latter part of the twentieth century, one influential view was that some kinds of values could influence science without causing problems, whereas other kinds of values needed to be carefully limited. For example, Ernan [McMullin \(1982\)](#) argued that “epistemic” values, which provided evidence that a theory was true, could appropriately influence scientific reasoning. In contrast, other values, the “non-epistemic” ones, did not provide evidence that a theory was true, and therefore, they should not influence scientific reasoning. Scholars have since added further nuance to this distinction. [Larry Laudan \(2004\)](#) and [Heather Douglas \(2013\)](#) argued that the category of “epistemic” values failed to distinguish between values that were genuinely truth indicative and those that were valued by scientists for other cognitive reasons, such as being pragmatically useful. Daniel Steel argued further that epistemic values could be distinguished into those that are intrinsic, meaning that exemplifying them “either constitutes an attainment of truth or is a necessary condition for a statement to be true,” as opposed to those that are extrinsic, meaning that they “promote the attainment of truth without themselves being indicators or requirements of truth” ([2010](#), 18).

Although the epistemic/non-epistemic distinction continues to be widely used, serious challenges have been leveled against the possibility of drawing a strict distinction of this sort ([Longino 1996](#); [Rooney 1992](#)). Part of the problem is that values that normally seem indicative of a theory’s truth (e.g., its coherence with other established theories) might not always be genuine indicators of its veracity, and values that do not normally seem like genuine indicators of a theory’s truth (e.g., its coherence with our considered ethical judgments) could sometimes help us arrive at more reliable theories. Another challenge is that so-called epistemic values typically need to be weighed and interpreted, and it is plausible that non-epistemic values typically play a role in that process. One response to this problem is to assert that there may be some cases in which values can be classified relatively cleanly as epistemic or non-epistemic, but there are other cases where values occupy a messy “borderlands” between the two ([Rooney 2017](#)). We have included a chapter on aesthetic values in the handbook (see [Ivanova 2026](#)), which illustrates the interesting philosophical work that can be done by studying values that occupy the borderlands between the epistemic and the non-epistemic.

In recent years, philosophers studying the role of values in science have argued that even “non-epistemic” values have legitimate roles to play in scientific reasoning. These philosophers typically claim to be rejecting a “value-free ideal” (VFI) that dominated late-twentieth-century philosophy of science. Based on her historical analysis of this time period, [Douglas \(2009\)](#) characterizes the VFI as the view that “the value judgments internal to science, involving the evaluation and acceptance of scientific results at the heart of the research process, are to be as free as humanly possible of all social and ethical values” (2009, 45). It is important to recognize, however, that this VFI does not completely exclude values from science. It makes a distinction between the “internal” aspects of science, which involve the core processes of scientific reasoning, and the “external” aspects of science, such as choosing research topics and making use of science for policy making. This VFI allows non-epistemic values to influence the external aspects of science, and it allows epistemic values to influence the internal aspects of science. Thus, Douglas notes that the VFI is actually a “somewhat inaccurate label” for what is really an “epistemic-values-only-in-scientific-inference” ideal ([2016](#), 3).

It is crucial to remember that even if one were to accept this VFI, one could still allow non-epistemic values to play very significant roles in science—e.g., guiding the directions of scientific inquiry, promoting inclusive and ethical scientific practices, and steering the communication of findings—which many of the chapters in this handbook discuss (see, e.g., [Beck et al. 2026](#); [Biddle 2026](#); [Bluhm 2026](#); [Douglas 2026](#); [Elabbar 2026](#); [Goldenberg 2026](#); [Gutiérrez Valderrama 2026](#); [Nieves Delgado 2026](#); [Resnik 2026](#); [Steel 2026](#)). In fact, philosophers have increasingly been arguing that the internal/external distinction does not hold up well under scrutiny and that value influences on the external aspects of science can end up influencing the internal aspects as well (see, e.g., [Elliott 2026](#); [Holman and Bruner 2017](#)).

Even though the VFI already allows for significant roles of values in science, [Douglas \(2000, 2009, 2021\)](#) famously argued that it was still ethically problematic. She noted that scientists, like all moral agents, have ethical responsibilities to avoid negligently or recklessly causing harm to others. On this basis, she argued that scientists ought to take into account the potential harmful effects of mistakes when making inductive inferences or engaging in other scientific activities (such as characterizing or interpreting data) that contribute to inductive inferences ([Douglas 2003](#)). For example, if scientists were studying a potentially deadly chemical, they would be morally required to consider the consequences of predicting the deadliness of the chemical incorrectly. If they concluded the chemical was deadly and it wasn't, then economically burdensome (and needless) regulations would occur based on their conclusion; if they concluded the chemical was not deadly and it was, then there would be (preventable) human deaths. Social and ethical values are important for balancing these consequences to the economy and public health.

While Douglas's "argument from inductive risk" (AIR) became the main fulcrum for overturning the VFI, a number of scholars formulated additional arguments against it (see [Elliott 2026](#)). For example, another challenge to the VFI is based on the idea that science incorporates multiple aims, including non-epistemic ones—e.g., promoting environmental sustainability or protecting public health (see [Lusk and Elliott 2024](#)). Therefore, when scientists are determining how to model phenomena or assess hypotheses, they need to take non-epistemic values into account (see, e.g., [Elliott and McKaughan 2014](#); [Fernández Pinto and Hicks 2019](#)). For example, Kristen [Intemann \(2015\)](#) has emphasized that when climate models are optimized to predict one kind of phenomenon (e.g., global temperature changes), they may not be as good at predicting other phenomena (e.g., local changes in precipitation), so scientists need to consider the non-epistemic aims of society when assessing these climate models. Another important challenge to the VFI is based on the notion that scientists employ concepts that incorporate both epistemic and non-epistemic elements (e.g., [Dupré 2007](#)). For example, Anna [Alexandrova \(2018\)](#) argues that a concept like "well-being" has this sort of "mixed" character, meaning that scientists need to take non-epistemic values into account when deciding how to define it in an appropriate manner. (For further discussion of this argument, see [Lewens \(2025\)](#) and [Risjord et al. \(2025\)](#); see also the chapters by [Anderson \(2026\)](#), [Elabbar \(2026\)](#), and [Larroulet Philippi \(2026\)](#)). In some respects, these arguments based on non-epistemic aims and concepts provide deeper critiques of the VFI than the AIR because they are not based on the presence of scientific uncertainty ([de Melo-Martín and Intemann 2016](#)). In contrast, the importance of values under the AIR diminishes as scientific uncertainty decreases because the potential for error also decreases ([Douglas 2000, 577; 2009, 96](#)).

Based on these arguments, it is now widely accepted that the VFI is untenable (see, e.g., [Biddle 2013](#); [Brown 2020](#); [Douglas 2009, 2016](#); [Elliott 2022b](#); [Intemann 2015](#); [Kitcher 2011](#); [Wilholt 2009](#); [Winsberg 2024](#)). For example, Douglas's AIR has been shown to be valid and defended as sound ([Brown 2024](#); [Brown and Stegenga 2023](#); [Havstad 2022](#)). Some authors, however, have resisted this conclusion (e.g., [Betz 2013](#); [Hudson 2021](#); [Lacey 2017](#)), and a number of philosophers have attempted to resurrect some form of the VFI (see, e.g., [Levy 2025](#); [Menon and Stegenga 2023](#); [Parker](#)

2024; Sheykh-Rezaee and Bikaraan-Behesht 2023; Sikorski 2024; Stamenkovic 2024). Perhaps surprisingly, this scholarship itself appears to be motivated by a mixture of epistemic and non-epistemic considerations (Ambrosj et al. 2023). For example, some of those seeking to limit the influences of non-epistemic values on scientific reasoning appeal to social concerns, such as the importance of maintaining science as a trustworthy source of information on which all members of society can rely (see, e.g., Betz 2013; Bright 2018; John 2015; Menon and Stegenga 2026). Yet, no matter how reasonable these motivations may appear, it has been argued that none of the current approaches to saving the VFI are successful (Brown 2024, 2026).

Nevertheless, these debates over the VFI are providing opportunities to sharpen our conceptual understanding of it (Elliott 2025b). To assess the cogency of efforts to salvage some form of the VFI, it is important to develop clear notions of what values are (Douglas and Elliott 2022; Elliott and Korf 2024; Hilligardt 2022; Winsberg 2024), what it is for scientific reasoning to be free of values (Ward 2021, 2026), and what is involved in holding onto an ideal (de Melo-Martín and Intemann 2016; Lusk 2025; Menon and Stegenga 2023). When debating formulations of the VFI, it is especially important to recognize that a scientific choice can be “value-laden” in at least four different ways: Values could serve as *motivating reasons* for the choice, they could serve as *justifying reasons* for the choice, they could play a *causal role* influencing the choice, or they could be *affected by* the choice (Ward 2021). Without clarifying these distinctions, it is easy for proponents and opponents of the VFI to talk past each other. For example, a proponent of the VFI might insist that it is possible for scientists to avoid consciously appealing to non-epistemic values (thereby excluding values as *motivating* reasons), whereas an opponent of the VFI might insist that non-epistemic values still ought to be taken into account (thereby including values as *justifying* reasons; see Ward 2021, 2026).

1.3 Managing values

Whatever one concludes about these debates regarding the VFI, it is clear that steps need to be taken to manage values in science appropriately. On one hand, the critics of the VFI need to show that non-epistemic values can be incorporated responsibly in scientific reasoning without threatening the integrity of science (Resnik and Elliott 2019). On the other hand, the defenders of the VFI are essentially proposing a particularly stringent form of value management—namely, excluding non-epistemic values completely (or at least *striving* to exclude non-epistemic values) from some aspects of scientific reasoning. Partly for this reason, figures like Brown (2024, 2026) and Lusk (2025) argue that it would be more productive to stop talking about the VFI and simply focus on discussing the best ways to manage values in science.

In fact, value management is an important issue no matter what kinds of values are at play. Historically, discussions of value management have typically focused on managing non-epistemic values, thereby implicitly suggesting that epistemic values do not require management. But this is a problematic assumption. As long as scientists need to make debated judgments about how to interpret and weigh values relative to one another (whether those values are classified as epistemic or non-epistemic), it raises the potential for reasonable and legitimate scientific disagreement. This, in turn, calls into question science’s character as a universal, authoritative source of neutral information on which everyone can rely (Betz 2013; Holman and Wilholt 2022; Lusk 2021). Therefore, steps need to be taken to manage these value-based scientific disagreements in socially responsible ways. For example, suppose a number of prominent climate scientists disagree about which climate models will generate more reliable predictions. Perhaps some scientists recommend one model because of its consistency with earlier models and because of its simplicity. Perhaps other scientists recommend a different model because they think it more accurately represents underlying features of the physical climate system. Even if all the scientists involved in the dispute think that they are appealing solely to

epistemic values, they still need to find ways to manage these value-based disagreements responsibly when they are making claims that could have a significant impact on society.

Value management, then, is one of the most important topics in the current literature on values and science. When [Douglas \(2009\)](#) first challenged the VFI, she suggested a management approach that distinguished between direct and indirect roles that values could play in the internal stages of scientific reasoning. She argued that values should not play a *direct* role in these stages (meaning they should not be treated as a form of evidence or as a reason in themselves for making a choice); rather, values should be limited to an *indirect* role (meaning that they should be used to assess whether the available lines of evidence/reasons are sufficient to draw a conclusion; [Douglas 2018](#), 3). Some figures have questioned the cogency of this distinction (e.g., [Elliott 2013](#); [Steel and Whyte 2012](#)), but it still captures a common intuition that values are different from evidence. Nevertheless, [Brown \(2020\)](#) has argued that this intuition rests on the problematic assumption that values do not have cognitive status. He challenges this assumption and argues that values can act as evidence under some circumstances as long as they do not predetermine the outcome of inquiry (see also [Anderson 2004](#)).

[Holman and Wilholt \(2022\)](#) recently divided the current approaches to value management into five main categories (see also [Elliott 2022b](#)), and they characterized questions about how to manage values in science as a “new demarcation problem.” In contrast to the old demarcation problem, which attempted to distinguish science from pseudoscience, the new demarcation problem (NDP) attempts to distinguish legitimate and illegitimate influences of values in science. An important theoretical question to consider moving forward is how to choose between value management approaches. Are there *a priori* reasons to think that one approach is typically better than the others? Or do different value management approaches work better in different situations? Some authors have even argued that the elements of multiple value management approaches can be integrated together ([Resnik and Elliott 2023](#); [Rolin and Koskinen 2022](#)). [Holman and Branch \(2026\)](#) in this handbook survey attempts to address the NDP.

One of the most important developments in the recent literature on managing values in science has been the turn toward a more “political” philosophy of science (see, e.g., [Badano 2025](#); [Hilligardt 2024](#); [Lusk 2021](#); [Scherz and Garcia-Portela 2025](#); [Schroeder 2021](#); [Ward and Creel 2024](#)). This scholarship builds on previous work by scholars from a number of different fields who have been striving to make science more responsive to public values (see, e.g., [Brown 2009](#); [Hess 2016](#); [Jasanoff 2017](#); [Pamuk 2021](#); [Sarewitz 2010](#); [Wynne 1996](#)). Within the philosophy of science, Philip [Kitcher \(2001, 2011\)](#) famously argued that science should be “well-ordered,” meaning that it should be directed in accordance with what an idealized group of deliberators would choose if they were informed both about relevant scientific facts and about the perspectives of all affected parties. In general, this kind of effort to integrate perspectives from political philosophy into value management is designed to address the worry that value-laden science cannot play a legitimate role in social decision making in a diverse society where people hold very different values. Proponents of a political philosophy of science contend that value-laden science can still have political legitimacy if the values underlying it are generated through appropriate political structures or processes, although this depends on broader normative judgments about what counts as a legitimate political process. Greater collaboration from political philosophers, who have not always taken the complexity of value-laden science into account ([Hamilton 2024](#)), could help move this scholarship forward. The chapters in [Part III](#) explore these issues in more detail.

1.4 Emerging issues

Studies of values and science have clearly made significant progress in clarifying the conceptual terrain and exploring how to manage value influences in science. These conceptual issues will likely continue to be important moving forward. One important area of contemporary scholarship is the

effort to distinguish different kinds of “epistemic risks” in addition to inductive risk (Biddle and Kukla 2017). For example, Stephanie Harvard and Eric Winsberg argue that scientists have to make value-laden decisions about how to handle the “representational risk” of adopting models that are inadequate to the purposes for which they were developed (see, e.g., Harvard and Winsberg 2022, 2026). Some scholars are also starting to discuss the “ontic risk” involved in drawing scientific conclusions that could affect the very phenomena under investigation (Ortmann 2025; Witteveen Under review). For example, consistently claiming that the climate crisis is a tragedy of the commons could make it even more likely to play out as a tragedy of the commons, whereas framing it in other ways could make it less likely to play out in this fashion (Ortmann and Veit 2023; Steel 2026).

Another important question is how to integrate an institutional and organizational perspective into the literature on values and science. When thinking about the topic of values and science, it can be tempting to focus on individual scientists and the values they bring to their work. This turns out to be a highly impoverished perspective (see Elabbar 2026; Koskinen and Rolin 2021; Ludwig and El-Hani 2025). As discussed above, Zina Ward (2021) has emphasized that choices in science can be value-laden not only in the sense that scientists are consciously motivated by values (i.e., beliefs or attitudes) but also in the sense that their choices affect values in the world (e.g., social justice, sustainability, public health). This opens up the opportunity to think about how institutions—the rules and norms that structure scientific practice (Jupille and Caporaso 2022)—influence scientific choices in ways that support some values and detract from other values (see Part IV).

These rules and norms are often generated and enforced by organizations, such as universities, scholarly societies, companies, contract research organizations, and standard-setting bodies. These organizations can exert powerful influences on research. For example, a wide-ranging interdisciplinary body of literature has recently documented the ways in which corporations have suppressed science that conflicts with their interests, funded questionable studies designed to produce desired results, and harassed scientists who present opposing perspectives (Fernández Pinto and Leuschner 2026; Holman and Elliott 2018; Michaels 2008; Oreskes and Conway 2010; White and Bero 2010). Some of these activities are obviously problematic, but in other cases industry exerts subtler value influences that deserve more nuanced analyses (Bravo 2026; Elliott 2014; Wilholt 2006). Comparing the characteristics of science produced by industry, academia, and nongovernmental organizations is likely to uncover important ways in which the organizational contexts in which science is produced—together with the explicit and implicit rules, guidelines, and standards associated with those contexts—serve particular values over others (Bravo and Elliott 2025).

Another important direction for future work is to explore the nuances of how value-laden choices permeate particular scientific activities or fields (see Parts V and VI). For example, there has been increasing attention to the role of values in specific aspects of science, such as modeling (e.g., Biddle and Winsberg 2010; Harvard and Winsberg 2022; Intemann 2015), measurement (e.g., Larroulet Philippi 2026; Ohnesorge 2025), classification (Brigandt 2022; Elliott 2009; Katikireddi and Valles 2015; Ludwig 2016), and communication (Franco 2019; John 2019; Kovaka 2021). Philosophers are also beginning to pay more attention to the relationships between different knowledge systems, including the roles that values play in those relationships (Kendig 2020; Koskinen and Rolin 2021; Ludwig and El-Hani 2020; Parke and Hikuroa 2024). In addition, there is growing recognition that scientific assessments designed to guide policy making on topics like climate change are value-laden in numerous ways and worthy of detailed scrutiny (see Elabbar 2026 and Whyte and Sherpa 2026). Virtually any scientific field incorporates subtle value judgments that merit further investigation (see Part VI).

Finally, as society struggles to deal with rampant misinformation and skepticism about science, a crucial role for philosophers will be to clarify how attention to values can help to address these problems. One avenue, as discussed earlier, is to clarify obvious abuses of science by those who operate in bad faith (see, e.g., McGarity and Wagner 2008; Michaels 2008; Oreskes and Conway 2010).

Scholars working on values and science can collaborate with those working in social epistemology and network epistemology to explore how misinformation spreads (see [Korf and O'Connor 2026](#)). In other cases, there are more subtle value differences that contribute to scientific disagreement and dissent against prevailing scientific views. In those cases, it can be helpful to clarify those value differences and explore whether or not they can be justified ([Biddle and Leuschner 2015](#); [de Melo-Martin and Intemann 2018](#); [Elliott 2023](#); [Goldenberg 2026](#); [Miller 2021](#)). In the face of these sorts of controversies, the values and science literature can also help to identify better approaches for science communication (see, e.g., [Branch 2026](#); [Halpern and Elliott 2022](#)) and more sophisticated perspectives on how to make science worthy of public trust (see, e.g., [Contessa 2023](#); [de Melo-Martin and Intemann 2018](#); [Intemann 2024](#); [Irzik and Kurtulmus 2019](#); [Metzen 2024](#); [Wilholt 2013](#)).

The six parts of the handbook are designed to provide a survey of this widespread literature. Some chapters focus on the theoretical background for contemporary debates about values and science, while others investigate the practical applications of this work. Some chapters provide reviews of existing literature, while others explore new and emerging issues. Many of the chapters draw connections to other areas of scholarship, such as political philosophy, science and technology studies, and research ethics. Together, the handbook's chapters illustrate the wide breadth and relevance of this research area.

1.5 Organization of the Handbook

1.5.1 Part I: Theoretical background on values and science

Part I explores the theoretical ideas that structure the field of values and science. It begins with some of the key historical ideas that provided a starting point for subsequent work. Intemann's chapter describes some of the major feminist critiques that began to challenge the notion that science is or should be value-free, and Kristina Rolin delves more deeply into one of those feminist approaches, namely, standpoint theory. One of the key issues raised by this feminist scholarship is the concept of objectivity and the extent to which objective science must be value-free, which Inkeri Koskinen considers in her chapter. Building on this historical background, Kevin Elliott's chapter describes six major arguments that have been developed against the VFI. In a complementary chapter, Matthew Brown evaluates a number of arguments for maintaining the VFI and concludes that they are unconvincing.

The remainder of **Part I** explores conceptual issues that continue to be important for thinking about values and science. Ivanova's chapter examines the role of aesthetic values in science, which is an important topic both because aesthetic values have received very limited serious attention in the existing literature and because aesthetic values open up new layers of complexity in the relationship between epistemic and non-epistemic values. Zina Ward discusses the concept of value-ladenness, which, as noted above, is crucial for promoting precise discussions about the VFI and about the roles of values in science more broadly. Finally, Stephanie Harvard and Eric Winsberg's chapter makes the case for distinguishing different kinds of "epistemic risks" in science in addition to inductive risk. In particular, they introduce the concept of "representational risk," which involves mistakes in building models rather than mistakes in making inductive inferences (see also [Biddle and Kukla 2017](#); [Harvard and Winsberg 2022](#)).

1.5.2 Part II: Managing values in science

Part II turns to the question of how to manage values in science appropriately. As mentioned previously, [Holman and Wilholt \(2022\)](#) argued in an influential article that the issue of value management should be framed as a "new demarcation problem" (NDP). Bennett Holman and T.Y. Branch's

contribution to the handbook introduces the contemporary literature on value management by reviewing the concept of the NDP and the recent literature on it.

One approach to managing values in science is to harken back to the ban on values that was central to the traditional VFI. In response to the arguments against the traditional VFI, proponents of these neo-VFIs utilize varying types of value prohibitions. The chapters by Wendy Parker and Tarun Menon and Jacob Stegenga provide two examples of this approach. Parker acknowledges that non-epistemic values are often relevant to scientific reasoning, but she suggests using an “epistemic projection” approach to value management that shifts value-laden decisions into the “problem selection” stage of scientific inquiry. For their part, Menon and Stegenga acknowledge that it might not be appropriate to completely remove non-epistemic values from scientific reasoning, but they argue that the VFI is still tenable as an action-guiding principle that calls for scientists to remove the influences of values on their inferences to the greatest extent possible (see also [Stegenga and Menon 2023](#)).

Another particularly prominent approach to managing values in science is what [Holman and Wilholt \(2022\)](#) called the “systemic” approach. Systemic strategies for managing values in science focus on creating an appropriate social structure that can critically assess value influences on science. Helen Longino’s critical contextual empiricism (CCE) is arguably the most influential version of a systemic approach (see also [Solomon 2001](#)). Kirstin Borgerson provides an overview of CCE in her chapter and considers major critiques of it. As Borgerson emphasizes, CCE calls for critical discussion of values and background assumptions in science from a variety of different perspectives. A critical question for this sort of systemic approach, however, is what range of different perspectives are necessary and appropriate in order to foster objectivity in science. In her contribution, Juliana Gutiérrez Valderrama argues that geographical diversity has received inadequate attention, and this has resulted in a global scientific endeavor that prioritizes the values of privileged geographical locations. Additionally, Abigail Nieves Delgado and her colleagues clarify that the pursuit of diversity, especially data diversity, can also have costs, including stereotyping and extractive practices (see also [Bluhm 2026](#)). Therefore, they argue that it is often better for scientists to aim for local, rather than global, forms of diversity.

An additional approach to value management is to foster transparency or openness about values so that people can make their own decisions about whether to accept the science or reject it as problematic (see, e.g., [Douglas 2008](#); [Elliott 2021](#); [Elliott 2022a](#); [Metzen 2024](#)). This approach resonates with current trends in science, such as the promotion of open science initiatives (see, e.g., [Leonelli 2023](#)). Nevertheless, transparency has limitations. In his chapter, Stephen John examines the strengths and weaknesses of transparency as a response to the value-ladenness of science. He argues that transparency initiatives can have costs for public audiences (e.g., promoting unwarranted distrust of science) and for the quality of research itself (e.g., incentivizing researchers to adopt sub-optimal research practices).

1.5.3 Part III: Values, science, and democracy

As discussed above, one of the most widely discussed approaches to managing values in science is to lean upon democratic principles and practices. The idea is that value-laden science can still play a legitimate role in the political sphere as long as the values underlying the science are democratically legitimate. This is such an influential idea, with so many issues and complexities to explore, that we have assigned a separate part of the handbook to this topic.

As discussed earlier, the democratization of science is not a new idea (see, e.g., [Jasanoff 2017](#); [Kitcher 2001](#); [Wynne 1996](#)), but Andrew Schroeder’s work on this topic has been particularly influential in the recent values and science literature (e.g., [Schroeder 2019, 2021](#); see also [Lusk 2021](#)). In his chapter for the handbook, Schroeder provides an overview of the key reasons for thinking that science should be responsive to the public, and he then highlights some of the important complications

that need to be addressed if this goal is to be pursued in a serious fashion. Karen Kovaka and Inmaculada de Melo-Martin probe even further into these challenges in their chapters. Kovaka argues that it is very difficult to measure the public's values in a manner that is both truly representative of their diversity and that can meaningfully guide the wide array of value judgments that permeate scientific research. De Melo-Martin explores how different stakeholders who are impacted by a particular area of science, such as pharmaceutical development and regulation under the auspices of the FDA, often have conflicting values.

Given these complexities of achieving democratized science, philosophers of science are increasingly turning to political theory in an effort to develop more sophisticated perspectives about how to relate values and science. In their chapter, Hannah Hilligardt and Torsten Wilholt use both ideal and non-ideal political theory to explore how science ought to be organized in order to fulfill the distinctive roles that it plays in democratic societies. Yiftach Fehige uses political theory to explore the additional question of how religious values should factor into decisions about science in democratic societies. This is an important contribution because religious values have received very little discussion in the recent values and science literature.

Another topic that has received very little previous attention in values and science is how to think about science in *nondemocratic* contexts. Karoliina Pulkkinen provides a much-needed opportunity to reflect on this neglected topic in her chapter for the handbook. She problematizes the assumption that value influences flow in a simple top-down fashion in nondemocratic contexts, she examines the nature of ideology and the moral ambiguity around “wishful speaking” as a way to navigate ideology, and she raises the question of whether value influences in science could be legitimized in ways other than through some form of democratic endorsement.

Finally, Heather Douglas's and Alexander Tolbert's chapters conclude the discussion of value management approaches that runs through Parts II and III. Noting that science is a source of power in modern society, Douglas argues that proper value management in science should also include considerations of social justice. She considers how justice can be promoted or jeopardized through four aspects of research: (1) Participation in science; (2) the treatment of human subjects in science; (3) interaction and engagement with social communities through research; and (4) agenda-setting for science. Alexander Tolbert then provides a detailed analysis of injustice in a particular area of science, namely, machine learning. He argues that the current literature on algorithmic fairness in machine learning reproduces racial disparities in society because it treats fairness as a technical problem and fails to grapple with background conditions of social injustice. Ultimately, Douglas's and Tolbert's chapters show that if social justice is to be achieved, it will stem not only from individual efforts by scientists but also through the creation of policies that guide scientific research toward the promotion of justice. Building on this point, the next part of the handbook explores the wide variety of institutions that influence how science impacts society.

1.5.4 Part IV: Values, science, institutions, and organizations

Philosophers of science have been arguing for years that more attention should be paid to the institutions and organizations that guide the practices of science in ways that support some values and challenge others (Biddle 2024; Douglas 2018; Elliott 2017; Fernández Pinto 2024). Nevertheless, there has still been a tendency in the literature to focus on the values that individual scientists or research groups bring to their work rather than to the institutions and organizations in which scientists operate. The chapters in this part examine the important relationships between values and scientific institutions and open up a number of important questions for future work.

The first group of chapters focus broadly on the organizations that fund and perform research. Jamie Shaw begins by focusing explicitly on an analysis of the values that public, private, and

philanthropic funders bring to science. Pedro Bravo then provides an overview of the wide variety of ways that industry funding can influence scientific research. This description is valuable because industry funding provides one of the few examples of an institutional or organizational context that has received sustained attention by those working on values and science (for an overview, see [Holman and Elliott 2018](#)). Manuela Fernández Pinto and Anna Leuschner build on Bravo's analysis by focusing specifically on the ways that companies (as well as other actors) harass scientists in order to discourage them from communicating unwanted messages. This phenomenon of epistemic intimidation and its effects on scientific practice is a crucial emerging topic that merits significant attention. Rebecca Korf and Cailin O'Connor extend the discussion in the previous two chapters by showing how network modeling can provide a richer understanding not only of industry influences on science but also of a wide array of other ways in which values influence the communal generation and communication of scientific knowledge. Finally, Rachel Ankeny and Emily Buddle problematize simple distinctions between the values of industry and the values of academia in their analysis of agricultural genomics research collaborations in Australia. Furthermore, they call for those working in values and science to adopt approaches that are more empirical and that consider a richer array of institutional factors beyond funding sources (e.g., standards, infrastructures, and intellectual property regimes; see also [Ankeny and Leonelli \(2016\)](#)).

In addition to the organizations that fund and perform research, some of the most important organizations that influence science are regulatory and policy-making bodies. One important question in this arena is how values influence the information that scientists provide to these organizations. This question is considered in two chapters that explore the values involved in formulating environmental assessment reports. Ahmad Elabbar examines how the assessment reports produced by the Intergovernmental Panel on Climate Change (IPCC) incorporate values through inductive risk considerations, visualizations, and procedures for information curation and synthesis. Kyle Whyte and Pasang Sherpa inquire into how the development of Indigenous-led assessment reports could generate scientific assessments that highlight different values. Another important question about policy-making bodies is how their science-related claims and recommendations are received by the public. Maya Goldenberg's chapter explores this question. She examines conflicts over vaccine policy and highlights how these debates, which are often framed as a conflict between science and politics, actually reflect a much more complex array of values. Finally, in the concluding chapter of this part, David Resnik explores a third relationship between science and policy-making bodies, namely, the ways in which they regulate the practice of science. Specifically, he examines the institutional structures that regulate dual-use biomedical research, which has become particularly pressing in the wake of the COVID-19 pandemic.

1.5.5 Part V: Values in scientific activities

Part V turns to an analysis of values in specific scientific activities. This part does not provide a comprehensive analysis of all the ways values arise in the practices of science; rather, it provides a sampling of different practices and the ways values are intertwined with them.

The first two chapters explore how values are involved in the nitty-gritty details of scientific data collection and analysis. Cristian Larroulet Philippi examines value considerations in the process of measurement. Perhaps because measurement appears so straightforward and factual, there has been comparatively little attention paid to measurement in the values and science literature. Larroulet Philippi, however, shows that this is a complex topic that merits more attention (see also [Ohnesorge \(2025\)](#)). Zinhle Mncube explores value considerations in deciding whether to include race as a predictor in medical guidelines and algorithms. Mncube argues that one can clarify the debate between proponents and opponents of using race as a predictor by attending to their differing approaches to

inductive risk (i.e., the potential to make false positive or false negative errors) and the values underlying those approaches.

The final four chapters of this part explore a broader array of scientific practices that are bound up with values. Charbel El-Hani and David Ludwig examine how different knowledge systems (e.g., academic, Indigenous, local) can be compared, contrasted, and integrated with each other. El-Hani and Ludwig highlight the dangers of oversimplifying either the similarities or the differences between knowledge systems, and they argue that one can avoid many of these dangers by focusing on the partial overlaps between different systems. Daniel Steel focuses on the role of values in framing issues at the intersection of science and policy. He uses climate change as a case study and explores the implications of framing it as a tragedy of the commons as opposed to an energy transition. T.Y. Branch discusses how different models for science communication reflect and reinforce different normative assumptions about the roles for values in science. Finally, Kristina Rolin examines research ethics education and considers how it can be enhanced by incorporating themes from the literature on science and values. In particular, she argues that scientists can better appreciate the importance of ethical norms and guidelines in science when they understand how those norms and guidelines support not only ethical and social goals but also epistemic goals.

1.5.6 Part VI: Values in specific sciences

Part VI concludes the handbook by considering how values are incorporated in the practice of specific sciences. Robyn Bluhm explores the tensions involved in trying to overcome the influences of androcentric values in the biomedical sciences. She shows that efforts to overcome androcentric values by incorporating more female animals in preclinical research can actually backfire and reinforce sex essentialism. Yasmin Haddad and Celso Neto focus on human genomics, arguing that there are good reasons to incorporate non-epistemic values in this area of research in order to lessen potential harms, respond to underdetermination and inductive risk, and guide the direction of research. Turning to the environmental sciences, Greg Lusk explores the many ways that values are embedded in climate modeling and also discusses how those value influences can be managed. Meanwhile, Jacalyn Beck and her colleagues provide their perspective as practicing conservation biologists about how value-laden frames and narratives shape the ways they study and communicate about complex biological systems. Turning to the social sciences, Elizabeth Anderson looks at value debates in classical economics in order to draw lessons for critiquing contemporary economics, especially with respect to its assumptions about how to measure human welfare. Finally, Justin Biddle considers the role of values in artificial intelligence research and governance, highlighting the importance of values for problem framing and operationalization; data quantity, quality, diversity, and access; modeling and validation; and deployment.

Our goal in this part, as in previous parts of the handbook, is obviously not to provide a comprehensive examination of values in all areas of science. Rather, we aim to consider enough different sciences to illustrate how value-laden dimensions can be uncovered in many different fields. Many of the chapters in other parts of the handbook provide additional illustrations of value dimensions in other areas of science (see, e.g., [Ankeny and Buddle 2026](#); [Mncube 2026](#); [Nieves Delgado et al. 2026](#)).

Some readers will notice that our examples in this part do not come from theoretical areas of chemistry or physics. More work is needed to probe the roles of values in these areas of science, but it is important to note that many of the arguments found throughout the handbook apply to these areas of science as well. For example, the chapter on aesthetic values ([Ivanova 2026](#)) includes a wide array of examples from the physical sciences. Values associated with who participates in science and what research agendas get funded are also highly relevant to these fields (see [Douglas 2026](#) and [Shaw 2026](#)). In fact, because the potential social benefits of these areas of science are not as obvious as in

other areas of science, researchers in these fields have to make particularly significant value-laden decisions about how to responsibly communicate their work in order to motivate funding and generate public interest without succumbing to inappropriate hype (Intemann 2022; see also Branch 2026 and Steel 2026). Along these lines, Kent Staley (2017) has pointed out that considerations of inductive risk are relevant even in a field like physics because researchers have to decide how much evidence to demand before announcing high-profile discoveries to the public.

1.6 Conclusion

The chapters in this handbook provide an accessible, forward-looking orientation to the literature on values and science. Besides providing background on some of the classic theoretical debates that have animated the field, the chapters illustrate the breadth of this topic and its intersections with other fields. Much insight has been gained by engaging with work in research ethics, biomedical ethics, science and technology studies, and political philosophy. Scholarship on concepts that integrate epistemology and ethics, such as epistemic injustice (Elliott 2025a; Gagné-Julien 2021; Grasswick 2017; Irzik and Kurtulmus 2024), pragmatic encroachment (Kim 2017; Miller 2014), coupled epistemic-ethical issues (Katikireddi and Valles 2015; Tuana 2013), and performativity (Ortmann 2025; Van Basshuysen et al. 2021), is likely to generate promising insights for values and science as well. In order to make this scholarship as impactful as possible, we encourage continuing collaborations between philosophers working in this area and practicing scientists, policy makers, and activist groups (see Ludwig and El-Hani 2025; Plaisance and Elliott 2021; Toomey 2024). We hope this handbook inspires a wide range of future scholarship on the ways in which ethics, values, and social considerations are embedded in the reasoning practices and the content of the sciences. In our view, these deep connections between the ethical and the epistemic, between the social context of science and the content of scientific claims, are what make the field of values and science so distinctive, exciting, and socially important.

Acknowledgments

We would like to thank Katharina Bernhard, Jesse Hamilton, Arnon Levy, Vincenzo Politi, and Zina Ward for their extremely helpful feedback on an earlier draft of this chapter. Kevin is also grateful for funding support he received through a Research Award from the Alexander von Humboldt Foundation and through the SOCRATES Center at Leibniz University, Hannover, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project 470816212/KFG43.

References

- Alexandrova, Anna. 2018. “Can the Science of Well-Being Be Objective?” *British Journal for the Philosophy of Science* 69: 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Ambrosj, Jacopo, Kris Dierickx, and Hugh Desmond. 2023. “The Value-Free Ideal of Science: A Useful Fiction? A Review of Non-epistemic Reasons for the Research Integrity Community.” *Science and Engineering Ethics* 29: 1. <https://doi.org/10.1007/s11948-022-00427-9>.
- Anderson, Elizabeth. 2004. “Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce.” *Hypatia* 19 (1): 1–24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>.
- Anderson, Elizabeth. 2026. “Values in Economics.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Ankeny, Rachel A., and Emily A. Buddle. 2026. “Valuing Partnerships in Agricultural Genomics Research: Using More Empirically Informed Philosophical Approaches to Advance the Science and Values Debates.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.

- Ankeny, Rachel A., and Sabina Leonelli. 2016. "Repertoires: A Post-Kuhnian Perspective on Scientific Change and Collaborative Research." *Studies in History and Philosophy of Science* 60: 18–28. <https://doi.org/10.1016/j.shpsa.2016.08.003>.
- Badano, Gabriele. 2025. "Public Reason, Values in Science, and the Shifting Boundaries of the Political Forum." *Philosophical Studies* 182: 1133–55. <https://doi.org/10.1007/s11098-025-02323-x>.
- Beck, Jacalyn M., Claire F. Hoffmann, Kelly Kapsar, and Robert A. Montgomery. 2026. "The Power of Value-Laden Framing: Examples from Wildlife Conservation Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Biddle, Justin B. 2013. "State of the Field: Transient Underdetermination and Values in Science." *Studies in History and Philosophy of Science Part A* 44 (1): 124–33. <https://doi.org/10.1016/j.shpsa.2012.09.003>.
- Biddle, Justin B. 2024. "Organizations and Values in Science and Technology." *Philosophy of Science* 91 (5): 1159–68. <https://doi.org/10.1017/psa.2023.145>.
- Biddle, Justin B. 2026. "Artificial Intelligence: Values, Governance, and Policy." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Biddle, Justin B., and Anna Leuschner. 2015. "Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?" *European Journal for Philosophy of Science* 5 (3): 261–78. <https://doi.org/10.1007/s13194-014-0101-x>.
- Biddle, Justin B., and Eric Winsberg. 2010. "Value Judgments and the Estimation of Uncertainty in Climate Modeling." In *New Waves in Philosophy of Science*, edited by P. D. Magnus and Jacob Busch. Palgrave Macmillan.
- Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies in Values in Science*, edited by Kevin Elliott and Ted Richards. Oxford University Press.
- Bluhm, Robyn. 2026. "'Every Cell Has a Sex': Sex Essentialism in Biomedical Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Branch, T.Y. 2026. "Values in Science and Science Communication Models." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Bravo, Pedro. 2026. "Values and Industry-Funded Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Bravo, Pedro, and Kevin C. Elliott. 2025. "The Promise and Perils of NGO-Supported Science." *Philosophy Compass* 20: e70043. <https://doi.org/10.1111/phc3.70043>.
- Brigandt, Ingo. 2022. "How to Philosophically Tackle Kinds without Talking about 'Natural Kinds.'" *Canadian Journal of Philosophy* 52 (3): 356–79. <https://doi.org/10.1017/can.2020.29>.
- Bright, Liam Kofi. 2018. "Du Bois' Democratic Defence of the Value Free Ideal." *Synthese* 195 (5): 2227–45. <https://doi.org/10.1007/s11229-017-1333-z>.
- Brown, Mark B. 2009. *Science in Democracy: Expertise, Institutions, and Representation*. MIT Press.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Brown, Matthew J. 2026. "Recent Arguments for the Ideal of Value-Free Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Brown, Matthew J., and Jacob Stegenga. 2023. "The Validity of the Argument from Inductive Risk." *Canadian Journal of Philosophy* 53 (2): 187–90. <https://doi.org/10.1017/can.2023.37>.
- Contessa, Gabriele. 2023. "It Takes a Village to Trust Science: Towards a (Thoroughly) Social Approach to Public Trust in Science." *Erkenntnis* 88 (7): 2941–966. <https://doi.org/10.1007/s10670-021-00485-8>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2016. "The Risk of Using Inductive Risk to Challenge the Value-Free Ideal." *Philosophy of Science* 83 (4): 500–520. <https://doi.org/10.1086/687259>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2018. *The Fight Against Doubt: How to Bridge the Gap Between Scientists and the Public*. Oxford University Press.
- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2003. "The Moral Responsibilities (Tensions between Autonomy and Responsibility)." *American Philosophical Quarterly* 40 (1): 59–68. <https://www.jstor.org/stable/20010097>.
- Douglas, Heather. 2008. "The Role of Values in Expert Reasoning." *Public Affairs Quarterly* 22 (1): 1–18. <https://www.jstor.org/stable/40441475>.

- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2013. "The Value of Cognitive Values." *Philosophy of Science* 80 (5): 796–806. <https://doi.org/10.1086/673716>.
- Douglas, Heather. 2016. "Values in Science." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.
- Douglas, Heather. 2018. "From Tapestry to Loom: Broadening the Perspective on Values in Science." *Philosophy, Theory, and Practice in Biology* 10 (8): 1–8. <https://doi.org/10.3998/ptpbio.16039257.0010.008>.
- Douglas, Heather. 2021. *The Rightful Place of Science: Science, Values, and Democracy: The 2016 Descartes Lectures*. Edited by Ted Richards. Consortium for Science, Policy & Outcomes.
- Douglas, Heather. 2026. "Science and Social Justice." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Douglas, Heather., and Kevin C. Elliott. 2022. "Addressing the Reproducibility Crisis: A Response to Hudson." *Journal of General Philosophy of Science* 53 (2): 201–9. <https://doi.org/10.1007/s10838-022-09606-5>.
- Dupré, John. 2007. "Fact and Value." In *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, Alison Wylie, and John Dupré. Oxford University Press.
- Elabbar, Ahmad. 2026. "Values and Assessment Reports on Climate Change." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Elliott, Kevin C. 2009. "The Ethical Significance of Language in the Environmental Sciences: Case Studies from Pollution Research." *Ethics, Place & Environment* 12: 157–73. <https://doi.org/10.1080/13668790902863382>.
- Elliott, Kevin C. 2013. "Douglas on Values: From Indirect Roles to Multiple Goals." *Studies in History and Philosophy of Science* 44 (3): 375–83. <https://doi.org/10.1016/j.shpsa.2013.06.003>.
- Elliott, Kevin C. 2014. "Financial Conflicts of Interest and Criteria for Research Credibility." *Erkenntnis* 79 (Suppl 5): 917–37. <https://doi.org/10.1007/s10670-013-9536-2>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2021. "The Value-Ladenness of Transparency in Science: Lessons from Lyme Disease." *Studies in History and Philosophy of Science* 88: 1–9. <https://doi.org/10.1016/j.shpsa.2021.03.008>.
- Elliott, Kevin C. 2022a. "A Taxonomy of Transparency in Science." *Canadian Journal of Philosophy* 52 (3): 342–55. <https://doi.org/10.1017/can.2020.21>.
- Elliott, Kevin C. 2022b. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin C. 2023. "Navigating Dissent by Managing Value Judgments: The Case of Lyme Disease." *Synthese* 202 (5): 134. <https://doi.org/10.1007/s11229-023-04368-z>.
- Elliott, Kevin C. 2025a. "Alleviating Epistemic Injustice in Environmental Health Research: Strategies from Science and Values." *Topoi*. <https://doi.org/10.1007/s11245-025-10268-3>.
- Elliott, Kevin C. 2025b. "Characterizing the Value-Free Ideal: From a Dichotomy to a Multiplicity." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw, and Carla Fehr. Springer.
- Elliott, Kevin C. 2026. "Arguments Against the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Elliott, Kevin C., and Daniel McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21. <https://doi.org/10.1086/674345>.
- Elliott, Kevin C., and Rebecca Korf. 2024. "Values in Science: What Are Values, Anyway?" *European Journal for Philosophy of Science* 14: 1–24. <https://doi.org/10.1007/s13194-024-00615-3>.
- Fernández Pinto, Manuela. 2024. "Pandemic Science and Commercial Values: An Institutional Account of Values in Science." *Philosophy of Science* 91 (5): 1139–48. <https://doi.org/10.1017/psa.2023.125>.
- Fernández Pinto, Manuela, and Anna Leuschner. 2026. "Epistemic Intimidation and Illegitimate Value-Influences in Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Fernández Pinto, Manuela, and Dan Hicks. 2019. "Legitimizing Values in Regulatory Science." *Environmental Health Perspectives* 127 (3): 035001. <https://pubmed.ncbi.nlm.nih.gov/30870036/>.
- Franco, Paul L. 2019. "Speech Act Theory and the Multiple Aims of Science." *Philosophy of Science* 86 (5): 1005–15. <https://doi.org/10.1086/705452>.
- Gagné-Julien, Anne-Marie. 2021. "Wrongful Medicalization and Epistemic Injustice in Psychiatry: The Case of Premenstrual Dysphoric Disorder." *European Journal of Analytic Philosophy* 17 (2): S4–36. <https://doi.org/10.31820/ejap.17.3.3>.
- Goldenberg, Maya J. 2026. "Controversies in Vaccine Policy." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.

- Grasswick, Heidi. 2017. "Epistemic Injustice in Science." In *The Routledge Handbook of Epistemic Injustice*, edited by Ian James Kidd, José Medina, and Gaile Pohlhaus, Jr. Routledge.
- Gutiérrez Valderrama, Juliana. 2026. "Values in Global Science and the Relevance of Geographic Diversity." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Halpern, Megan K., and Kevin C. Elliott. 2022. "Science as Experience: A Deweyan Model of Science Communication." *Perspectives on Science* 30 (4): 621–56. https://doi.org/10.1162/posc_a_00398
- Hamilton, Jesse. 2024. "Public Reason, Values in Science, and Democracy." PhD diss., University of Pennsylvania. <https://repository.upenn.edu/handle/20.500.14332/60883>.
- Harris, Richard. 2017. *Rigor Mortis: How Sloppy Science Creates Worthless Cures, Crushes Hope, and Wastes Billions*. Basic Books.
- Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32 (1): 1–31.
- Harvard, Stephanie, and Eric Winsberg. 2026. "Representational Risk Revisited." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Harvard, Stephanie, Eric Winsberg, John Symons, and Amin Adibi. 2021. "Value Judgments in a COVID-19 Vaccination Model: A Case Study in the Need for Public Involvement in Health-Oriented Modelling." *Social Science & Medicine* 286: 114323. <https://doi.org/10.1016/j.socscimed.2021.114323>.
- Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Hess, David J. 2016. *Undone Science: Social Movements, Mobilized Publics, and Industrial Transitions*. MIT Press.
- Hillgardt, Hannah. 2022. "Looking Beyond Values: The Legitimacy of Social Perspectives, Opinions and Interests in Science." *European Journal for Philosophy of Science* 12 (4): 1–20. <https://doi.org/10.1007/s13194-022-00490-w>.
- Hillgardt, Hannah. 2024. "Science as Public Service." *European Journal for Philosophy of Science* 14: 45. <https://doi.org/10.1007/s13194-024-00607-3>.
- Holman, Bennett, and Justin Bruner. 2017. "Experimentation by Industrial Selection." *Philosophy of Science* 84 (5): 1008–19. <https://doi.org/10.1086/694037>.
- Holman, Bennett, and Kevin C. Elliott. 2018. "The Promise and Perils of Industry-Funded Science." *Philosophy Compass* 13 (11): e12544. <https://doi.org/10.1111/phc3.12544>.
- Holman, Bennett, and T.Y. Branch. 2026. "Reflecting on Responses to the New Demarcation Problem." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Hudson, Robert. 2021. "Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis." *Journal for General Philosophy of Science* 52 (3): 389–405. <https://doi.org/10.1007/s10838-020-09548-w>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5: 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Intemann, Kristen. 2022. "Understanding the Problem of 'hype': Exaggeration, Values, and Trust in Science." *Canadian Journal of Philosophy* 52 (3): 279–94. <https://doi.org/10.1017/can.2020.45>.
- Intemann, Kristen. 2024. "Value Transparency and Promoting Warranted Trust in Science Communication." *Synthese* 203 (2): 42. <https://doi.org/10.1007/s11229-023-04471-1>.
- Intemann, Kristen. 2026. "Feminist Challenges to the Value-Free Ideal of Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Irzik, Gürol, and Faik Kurtulmus. 2019. "What Is Epistemic Public Trust in Science?" *The British Journal for the Philosophy of Science* 70 (4): 114566. <https://doi.org/10.1093/bjps/axy007>.
- Irzik, Gürol, and Faik Kurtulmus. 2024. "Distributive Epistemic Justice in Science." *The British Journal for the Philosophy of Science* 75 (2): 325–45. <https://doi.org/10.1086/715351>.
- Ivonova, Milena. 2026. "Aesthetic Values in Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Jasanoff, Sheila. 2017. "Science and Democracy." In *The Handbook of Science and Technology Studies*, edited by Ulrike Felt, Rayvon Fouche, Clark A. Miler, and Laurel Smith-Doerr. MIT Press.
- John, Stephen. 2015. "Inductive Risk and the Contexts of Communication." *Synthese* 192 (1): 79–96. <https://doi.org/10.1007/s11229-014-0554-7>.
- John, Stephen. 2019. "Science, Truth and Dictatorship: Wishful Thinking or Wishful Speaking?" *Studies in History and Philosophy of Science Part A* 78: 64–72. <https://doi.org/10.1016/j.shpsa.2018.12.003>.

- Jupille, Joseph H., and James A Caporaso. 2022. *Theories of Institutions*. Cambridge University Press.
- Katikireddi, Srinivasa V., and Sean A. Valles. 2015. "Coupled Ethical-Epistemic Analysis of Public Health Research and Practice: Categorizing Variables to Improve Population Health and Equity." *American Journal of Public Health* 105: e36–42.
- Kendig, Catherine. 2020. "Ontology and Values Anchor Indigenous and Grey Nomenclatures: A Case Study in Lichen Naming Practices among the Sami, Sherpa, Scots, and Okanagan." *Studies in History and Philosophy of Science Part C* 84: 101340. <https://doi.org/10.1016/j.shpsc.2020.101340>.
- Kim, Brian. 2017. "Pragmatic Encroachment in Epistemology." *Philosophy Compass* 12 (5): e12415. <https://doi.org/10.1111/phc3.12415>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Korf, Rebecca, and Cailin O'Connor. 2026. "Networks and Values." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Koskinen, Inkeri, and Kristina Rolin. 2021. "Structural Epistemic (In)justice in Global Contexts." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo. Routledge.
- Koskinen, Inkeri, and Kristina Rolin. 2022. "Distinguishing between Legitimate and Illegitimate Roles for Values in Transdisciplinary Research." *Studies in History and Philosophy of Science* 91: 191–98. <https://doi.org/10.1016/j.shpsa.2021.12.001>.
- Kourany, Janet A. 2010. *Philosophy of Science After Feminism*. Oxford University Press.
- Kovaka, Karen. 2021. "Climate Change Denial and Beliefs about Science." *Synthese* 198 (3): 2355–74. <https://doi.org/10.1007/s11229-019-02210-z>.
- Kuhn, Thomas. 1977. "Objectivity, Value Judgement, and Theory Choice." In *The Essential Tension*. University of Chicago Press.
- Lacey, Hugh. 2017. "Distinguishing between Cognitive and Social Values. In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Larroulet Philippi, Cristian. 2026. "Values and Measurement." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Laudan, Larry. 2004. "The Epistemic, the Cognitive, and the Social." In *Science, Values, and Objectivity*, edited by Peter Machamer and Gereon Wolters. University of Pittsburgh Press.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press.
- Levy, Arnon. 2025. "A (Consequence-Oriented) Critique of the Argument from Inductive Risk." *Philosophical Studies* 182: 2079–99. <https://doi.org/10.1007/s11098-025-02319-7>.
- Lewens, Tim. 2025. "Towards a Direct Role for Values in the Heart of Science." *Philosophy of Science* 92: 850–67. <https://doi.org/10.1017/psa.2025.18>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Longino, Helen E. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson. Kluwer.
- Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.
- Ludwig, David. 2016. "Ontological Choices and the Value-Free Ideal." *Erkenntnis* 81 (6): 1253–72. <https://doi.org/10.1007/s10670-015-9793-3>.
- Ludwig, David, and Charbel N. El-Hani. 2020. "Philosophy of Ethnobiology: Understanding Knowledge Integration and Its Limitations." *Journal of Ethnobiology* 40 (1): 3–20. <https://doi.org/10.2993/0278-0771-40.1.3>.
- Ludwig, David, and Charbel N. El-Hani. 2025. *Transformative Transdisciplinarity: An Introduction to Community-Based Philosophy*. Oxford University Press.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Lusk, Greg. 2025. "Science, Values, and Action Guidance: Can We Stop Talking about the Value Free Ideal?" In *Values, Pluralism, and Pragmatism*, edited by Jonathan Y. Tsou, Jamie Shaw, and Carla Fehr. Springer.
- Lusk, Greg, and Kevin Elliott. 2024. "Against Global Aims for Science: Values, Epistemic Priority, and a Local Aims Approach." *Synthese* 204 (2): 47. <https://doi.org/10.1007/s11229-024-04633-9>.
- Macedo, Stephen, and Frances Lee. 2025. *In Covid's Wake: How Our Politics Failed Us*. Princeton University Press.
- McGarity, Thomas O., and Wendy E. Wagner. 2008. *Bending Science: How Special Interests Corrupt Public Health Research*. Harvard University Press.

- McMullin, Ernan. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (2): 2–28. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409>.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom Is Worth Pursuing." *European Journal for Philosophy of Science* 13 (4): 48. <https://doi.org/10.1007/s13194-023-00552-7>.
- Menon, Tarun, and Jacob Stegenga. 2026. "Democracy, Consensus, and the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Metzen, Hanna. 2024. "Trust in Scientific Expertise and the Varying Demands of Value Transparency." *Canadian Journal of Philosophy* 54 (5): 364–79. <https://doi.org/10.1017/can.2025.10004>.
- Michaels, David. 2008. *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford University Press.
- Miller, Boaz. 2014. "Science, Values, and Pragmatic Encroachment on Knowledge." *European Journal for Philosophy of Science* 4 (2): 253–70. <https://doi.org/10.1007/s13194-014-0087-4>.
- Miller, Boaz. 2021. "When Is Scientific Dissent Epistemically Inappropriate?" *Philosophy of Science* 88 (5): 918–28. <https://doi.org/10.1086/714854>.
- Mncube, Zinhle. 2026. "Science, Values, and Race Correction." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Nieves Delgado, Abigail, Jan Baedke, and Ann-Christin Fisher. 2026. "The Limits of Diversity in Science: The Case of Human Microbiome Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- O'Connor, Cailin, and James Owen Weatherall. 2019. *The Misinformation Age: How False Beliefs Spread*. Yale University Press.
- Ohnesorge, Miguel. 2025. "We Should Not Align Quantitative Measures with Stakeholder Values." *Philosophy of Science*: 1–18. <https://doi.org/10.1017/psa.2025.10154>.
- Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Ortmann, Jakob. 2025. "Performative Paternalism." *European Journal for Philosophy of Science* 15 (2): 25. <https://doi.org/10.1007/s13194-025-00651-7>.
- Ortmann, Jakob, and Walter Veit. 2023. "Theory Roulette: Choosing that Climate Change is not a Tragedy of the Commons." *Environmental Values* 32 (1): 65–89. <https://doi.org/10.3197/096327122X16452897197784>.
- Pamuk, Zeynep. 2021. *Politics and Expertise: How to Use Science in a Democratic Society*. Princeton University Press.
- Parke, Emily C., and Daniel Hikuroa. 2024. "Against Defending Science: Asking Better Questions about Indigenous Knowledge and Science." *Philosophy of Science* 91 (5): 1210–1220. <https://doi.org/10.1017/psa.2023.146>.
- Parker, Wendy S. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36. <https://doi.org/10.1017/psa.2023.107>.
- Plaisance, Kathryn S., and Kevin C. Elliott. 2021. "A Framework for Analyzing Broadly Engaged Philosophy of Science." *Philosophy of Science* 88 (4): 594–615. <https://doi.org/10.1086/713891>.
- Resnik, David B. 2026. "Values and Dual Use Biomedical Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Resnik, David B., and Kevin C. Elliott. 2019. "Value-Entanglement and the Integrity of Scientific Research." *Studies in History and Philosophy of Science Part A* 75: 1–11. <https://doi.org/10.1016/j.shpsa.2018.12.011>.
- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54 (2): 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Risjord, Mark, Kareem Khalifa, and Jared Millson. 2025. "Thick Concepts and Impartiality." *Philosophy of Science*. <https://doi.org/10.1017/psa.2025.10114>.
- Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1): 13–22. <https://doi.org/10.1086/psaprocbienmeetp.1992.1.192740>.
- Rooney, Phyllis. 2017. "The Borderlands between Epistemic and Non-Epistemic Values." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Sarewitz, Daniel. 2010. *Frontiers of Illusion: Science, Technology, and the Politics of Progress*. Temple University Press.
- Scherz, Antoinette, and Laura García-Portela. 2025. "Climate Scientists as Trustees in Public Reason: The Legitimacy of Political Institutions amid Non-Epistemic Values." *Critical Review of International Social and Political Philosophy*: 1–26. <https://doi.org/10.1080/13698230.2025.2499360>.
- Schroeder, S. Andrew. 2019. "Which Values Should Be Built into Economic Measures?" *Economics & Philosophy* 35 (3): 521–36. <https://doi.org/10.1017/S0266267118000317>.

- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Sheykh-Rezaee, Hossein, and Hamed Bikaraan-Behesht. 2023. "Value-Free Ideal Is an Epistemic Ideal: An Objection to the Argument from Inductive Risk." *Principia: An International Journal of Epistemology* 27 (1): 1.
- Sikorski, Michał. 2024. "Values, Bias, and Replicability." *Synthese* 203: 164. <https://doi.org/10.1007/s11229-024-04573-4>.
- Solomon, Miriam. 2001. *Social Empiricism*. MIT Press.
- Staley, Kent. 2017. "Decisions, Decisions: Inductive Risk and the Higgs Boson." In *Exploring Inductive Risk: Case Studies in Values in Science*, edited by Kevin Elliott and Ted Richards. Oxford University Press.
- Stamenkovic, Philippe. 2024. "Straightening the 'Value-Laden Turn': Minimising the Influence of Extra-Scientific Values in Science." *Synthese* 203: 20. <https://doi.org/10.1007/s11229-023-04446-2>.
- Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. <https://doi.org/10.1086/650206>.
- Steel, Daniel. 2026. "Tragedy or Transition? How Science and Values Matter for Climate Change Frames." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Steel, Daniel, and Kyle Whyte. 2012. "Environmental Justice, Values, and Scientific Expertise." *Kennedy Institute of Ethics Journal* 22 (2): 163–82. <https://doi.org/10.1353/ken.2012.0010>.
- Stegenga, Jacob, and Tarun Menon. 2023. "The Difference-to-Inference Model for Values in Science." *Res Philosophica* 100 (4): 423–47. <https://doi.org/10.5840/resphilosophica2023928102>.
- Toomey, Anne Helen. 2024. *Science with Impact: How to Engage People, Change Practice, and Influence Policy*. Island Press.
- Tuana, Nancy. 2013. "Embedding Philosophers in the Practices of Science: Bringing Humanities to the Sciences." *Synthese* 190 (11): 1955–73. <https://doi.org/10.1007/s11229-012-0171-2>.
- Van Basshuysen, Philippe, Lucie White, Donal Khosrowi, and Mathias Frisch. 2021. "Three Ways in Which Pandemic Models Perform a Pandemic." *Erasmus Journal for Philosophy and Economics* 14 (1): 110–27.
- Ward, Zina B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Ward, Zina B. 2026. "What Does It Mean to Say that Science Is Value-Laden?" In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Ward, Zina B., and Kathleen A. Creel. 2024. "To Hedge or Not to Hedge: Scientific Claims and Public Justification." *Philosophy of Science* 91 (4): 990–1008. <https://doi.org/10.1017/psa.2024.17>.
- White, Jenny, and Lisa A. Bero. 2010. "Corporate Manipulation of Research: Strategies Are Similar Across Five Industries." *Stanford Law & Policy Review* 21: 105.
- Whyte, Kyle, and Pasang Yangjee Sherpa. 2026. "Indigenous Peoples' Leadership in Environmental Science Assessments." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Wilholt, Torsten. 2006. "Design Rules: Industrial Research and Epistemic Merit." *Philosophy of Science* 73 (1): 66–89. <https://doi.org/10.1086/510175>.
- Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science* 40: 92–101. <https://doi.org/10.1016/j.shpsa.2008.12.005>.
- Wilholt, Torsten. 2013. "Epistemic Trust in Science." *The British Journal for the Philosophy of Science* 64 (2): 233–53. <https://doi.org/10.1093/bjps/axs007>.
- Winsberg, Eric. 2024. "Managing Values in Science: A Return to Decision Theory." *Kennedy Institute of Ethics Journal* 34 (4): 389–418. <https://dx.doi.org/10.1353/ken.2024.a965814>.
- Witteveen, Joeri. Under review. "From Inductive Risk to Ontic Risk."
- Wynne, Brian. 1996. "May the Sheep Safely Graze? A Reflexive View of the Expert-Lay Knowledge Divide." In *Risk, Environment and Modernity: Towards a New Ecology*, edited by Scott M. Lash, Bronislaw Szerszynski, and Brian Wynne. SAGE Publications.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART I

Theoretical Background on Values and Science



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

2

FEMINIST CHALLENGES TO THE VALUE-FREE IDEAL OF SCIENCE

Kristen K. Intemann

2.1 Introduction

Feminist scholars played an early and important role in arguing that values play a variety of roles in science (Bleier 1978; Collins 1986; Fausto-Sterling 1985; Haraway 1985; Harding 1986; Hrdy 1981; Hubbard 1982; Longino 1987, 1990; Nelson 1990; Okruhlik 1994; Smith 1987; Wylie et al. 1989). At the time, much of what they argued was perceived to be implausible by both scientists and philosophers of science precisely because it was viewed as inconsistent with the “value-free ideal” (VFI) of science (Gross 1994; Haack 1992). Broadly speaking, the VFI holds that science ought to be free of non-epistemic values (including social, ethical, or political values). Yet as Kevin Elliott (2025) has recently pointed out, there may be different conceptions of this ideal depending on how “values” are understood, what is meant by “freedom,” and how “ideal” is conceived. When feminists challenged the VFI, what were they challenging? This chapter will show that feminists had a distinct set of aims for engaging with science that ultimately led to several arguments that political values could play important roles in science. While these arguments did undermine the most widely held version of the VFI, the main target of their criticism was a set of conceptions of scientific objectivity that motivated the VFI. Indeed, a central contribution of feminists has been to provide alternative accounts of objectivity that do not require value-freedom. Some of these accounts of objectivity might be compatible with some alternative version of the VFI, though whether these would be helpful ideals may depend on what they are supposed to protect or regulate. I will begin by identifying three aims feminists had for engaging with science and philosophy of science. I will then consider how feminist work in pursuit of these aims was seen as challenging the VFI. I then show how this work led to alternative conceptions of scientific objectivity, some of which might be compatible with revised versions of the VFI.

2.2 The science question(s) in feminism

Some have found it puzzling why feminist scholars concerned themselves with science when, on the surface, feminism and science appear to involve different areas of inquiry concerned with distinct aims (Gross 1994; Gross and Levitt 1994; Haack 1992). Feminism is a political movement challenging systems of oppression to make the world more just, while science aims to discover the way the

world is. Yet feminist political goals inevitably required philosophical engagement with, and scrutiny of, science for several reasons.

First, feminist political aims necessitated engaging with science, because it was often used to justify or explain existing inequalities. Going back to the first wave feminist suffrage movements, feminist political aims and policy recommendations were often challenged on the basis that they were inconsistent with scientific evidence. For example, various sex differences in the brain were claimed to result in different cognitive abilities (Gould 1996). In the early twentieth century, several prominent scientists and physicians not only published research but also wrote opinion pieces, letters to the editor, and made public lectures to explain why, scientifically, women should not be allowed to vote (Jorgensen-Earp and Jorgensen 2016). Blood flow differences in the brain were claimed to make women less capable of original thought, less suited for complex intellectual work, and unlikely to exercise calm judgment (Maney 2016). In the later twentieth century, even as leading scientific theories changed, new theories were used for the same purposes. Structural differences in the brain, sex differences in IQ, or sex chromosomes were used to justify social inequalities, including differences in education, underrepresentation in certain professions, or income disparities (Bluhm 2020; Fausto-Sterling 1985; Gould 1996; Jordan-Young 2010; Richardson 2013). Therefore, feminists were motivated to critically examine the science that had been used to justify existing inequalities. Yet their criticisms were often discounted precisely because science was thought to provide “objective facts” about sex differences that could not be reasonably challenged. Consequently, feminists began to not only examine particular studies purporting to show sex differences but also question philosophical conceptions about the nature of science, objectivity, and scientific reasoning. Were the results of scientific inquiry “objective facts” or might it be that accepted theories sometimes reflect the social values in which science occurs and/or the interests of those practicing science? In particular, they were motivated to consider how gendered norms, assumptions, and values might influence scientific reasoning.

Second, feminists were interested in engaging with science because it was an area in which women had historically been excluded and continued to be underrepresented (Rossiter 1982; Valian 1997). The second wave feminist movement of the 1960s–1980s focused on women’s rights, including equal opportunities for education, work, and pay. Understanding the kinds of challenges that women faced in science, technology, engineering, and math fields, throughout their education and professional pipeline, was important to understanding the obstacles that women and other represented groups faced in education and the workplace. Yet in investigating why women were underrepresented and how best to address those challenges, feminists also had to consider why it mattered. Certainly, it mattered ethically—that is, it would be wrong to exclude women from STEM education or career opportunities on the basis of gender. But feminists also wondered whether it might make a difference epistemically, or whether the inclusion of women could produce better science. Indeed, if gendered norms or sexist assumptions did influence some areas of study, might more diversity among scientists potentially address that problem in some way? If practicing scientists were more representative of society, would that change either how knowledge is produced or what is known (Harding 1986, 21)?

Finally, this led to a third feminist motivation for engaging with science. Put simply, advancing feminist political aims requires knowledge of the way the world is. That is, challenging systems of oppression necessitates understanding how the world operates. It requires understanding certain categories, such as sex, gender, sexuality, and race, that have been thought to involve differences and create disparities. It also involves understanding the ways in which inequalities are produced and maintained through our language, biology, psychology, and institutions (including political, educational, and scientific institutions). Moreover, feminist political objectives required knowledge about which policies or actions or practices might help in achieving greater equality. In other words, advancing feminist aims required knowledge and feminists recognized that science might be a powerful

way to produce such knowledge. Thus, many feminist philosophers of science, while sometimes critical of science, recognized the power of knowledge production and were motivated to consider the emancipatory potential of scientific inquiry (e.g., [Harding 1986](#); [Kourany 2003, 2010](#); [Longino 1990](#)).

These three motivations for engaging with science and philosophy of science (identifying and evaluating gendered norms and values in science, understanding the underrepresentation of women in STEM, and considering the emancipatory potential of science) were feminist in the sense that they arose out of broad feminist political aims. This set of motivations also led feminists to challenge a set of epistemological assumptions that supported the VFI. The next section will examine how specifically feminist engagement with science led them to ultimately question the roles of values in science and the nature of scientific objectivity.

2.3 The VFI and objectivity

The view that science ought to be value-free has shifted in meaning over time to be utilized for different purposes ([Douglas and Branch 2024](#); [Proctor 1991](#)). When feminists began to criticize research on sex differences, or question what could be inferred from biological differences, they were often dismissed on the basis that such criticism amounts to an attempt to “politicize” science (see e.g., [Fine 2018, 2022](#)). Thus, feminist critiques of science often ran up against a crude version of the VFI, which might be understood as the claim that science should be completely free of political values at every stage, including motivations for inquiry. On this version, science (or at least “pure” science) should be inquiry for its own sake, rather than in the service of any political aims. Of course, it is not clear that scientific research into sex differences was *not* politically motivated (given that it was often used to justify particular political policies), but feminists were forced to question whether all politically motivated research was problematic (in virtue of being political). The view that political or other non-epistemic values should not motivate inquiry seems to reflect a particular conception of objectivity that Heather Douglas has referred to as “detached objectivity” ([Douglas 2004](#), 549). The idea is that researchers should be “detached” from any political motivations, to keep one from wanting a particular outcome of inquiry, or be so politically driven so as to fail to see evidence or explanations that might lead to an alternative outcome. As Douglas argues, this conception of objectivity, however, often rapidly expanded into the broader claim that objectivity requires science to be completely free of values ([Douglas 2004](#), 459).

This crude version of the VFI—that values should have no role in science (including motivations for what is studied, examined, or criticized)—was quickly recognized to be implausible. Philosophers acknowledged that social, ethical, and political values can indeed sometimes play a role in determining whether or why something is studied, how research questions are formulated, how research might be applied to implementing new innovations or science-based policies, or whether certain kinds of methodologies would be ethical ([Douglas 2009](#); [Kitcher 2001](#); [Kourany 2010](#); [McMullin 1982](#)). Proponents of the VFI thus refined the ideal in ways to refine by appealing to a distinction often attributed to Hans [Reichenbach \(1938\)](#) between the “context of discovery” or the factors that influence certain aspects of the world to be investigated or theories to be formulated, and the “context of justification” or the context of assessing evidence and evaluating which theories ought to be accepted (see also [Schickore 2022](#)). As Matt [Brown \(2024\)](#) recently puts it, the VFI is the claim that

only scientific or ‘epistemic’ values can influence scientific reasoning or inference, while the only place for other values, including social and ethical values, should be in external aspects of science, such as choice of research projects or decisions about acceptable methods. (112)

This is the version of the VFI that has been largely used in the science and values literature over the past several decades (e.g., [Brown 2020](#); [Douglas 2009](#); [Elliott 2017](#); [McMullin 1982](#)).

Yet while the VFI was refined, the underlying goal was the same: To protect the rationality and objectivity of science. By narrowing the focus to the value-freedom to scientific reasoning and inference, the “heart” of science could be defended as rational and objective. (e.g., [McMullin 1982](#); [Nagel 1979](#); [Ruse 1984](#)).¹ It would be more precise, however, to say that the VFI protected a particular set of conceptions of objectivity (which were often viewed as interconnected). Specifically, the VFI can be thought to protect the objectivity of knowledge claims, methods, and knowers ([Crasnow and Intemann 2024](#)).² The VFI has been thought as needed to protect the objectivity of knowledge claims insofar as objective scientific knowledge is understood as “knowledge that bears no trace of the knower—knowledge unmarked by prejudice or skill, fantasy or judgment, wishing or strivings” ([Daston and Gallison 2007](#), 17). Some wanted to establish that the products of science—particular scientific claims or theories—were objective, in the sense that they corresponded to reality, and not merely “myths” made up by humans ([Lloyd 1995](#); [Ruse 1984](#)). In addition, the products of science were claimed to be objective insofar as they were produced by methods that were reliable in producing true or empirically adequate claims. Thus, the VFI was also thought to protect the objectivity of methods whereby, if properly employed, they would achieve the same results regardless of who employs them ([Crasnow and Intemann 2024](#)). That is to say, objective methods also “bear no trace of the knower.” Such methods could not include non-epistemic values because (it was often assumed) those sorts of values are not truth conducive and would be mere subjective preferences. Finally, the objectivity of inquirers was thought to safeguard the proper use of objective methods ([Crasnow and Intemann 2024](#)). That is, scientists should strive to not allow non-epistemic values to influence their use of scientific methods, their interpretation of data, or their inferences or reasoning.

The motivation for the refined version of the VFI, then, is the result of a set of interrelated conceptions of objectivity that were assumed to protect the rationality of science. Scientists should prevent the influence of non-epistemic values because they will be more likely to investigate, evaluate, and assess hypotheses using objective methods, and this will result in objective knowledge that captures what Elizabeth [Lloyd \(1995, 353\)](#) has referred to as the “really real.”

These interrelated conceptions of objectivity also perhaps rendered feminist critiques of science unintelligible. For those who held one or more of these conceptions, feminist critiques of particular scientific studies seemed misguided precisely because they were criticizing “objective facts,” for what were assumed to be non-epistemic reasons. Similarly, this conception of “objective scientists” made feminist calls for greater diversity in science non-sensical. That is, even if one believes it would be unjust to keep certain groups from participating in education or science for political reasons, it could not make any difference epistemically because the gender or race or class of the scientist was irrelevant to the production of scientific knowledge. The job of objective scientists was to put aside any personal or political values in scientific reasoning. Indeed, because one’s social features were believed to be epistemically irrelevant to scientific work, this also made it more dubious that it even mattered ethically if certain groups were underrepresented. That is, if such features are not relevant to what makes one qualified for scientific work, then it would be unjust to “discriminate” on that basis or take gender into account at all. Finally, if scientists must strive to ensure they are not being influenced by values in order to be objective and employ methods correctly to achieve objective facts, then the idea that science could also be used for emancipatory purposes was also misguided. Objective methods aim only to uncover truths about the world, regardless of their implications for particular political aims (such as equality or justice).

As we shall see, feminists provided several arguments and examples, demonstrating that values play necessary and important roles in science in a variety of ways. Yet the breadth of feminist

contributions here is best understood as part of a larger critique of how we understand scientific objectivity that the VFI was assumed to support.

2.4 The value question(s) in science

Particularly as more women and feminists began to enter scientific fields in larger numbers in the 1970s–1990s, they began to reveal ways in which sexist, androcentric, heterosexist, racist, and classist values, norms, and assumptions *had* influenced scientific theorizing, observations, and reasoning in a variety of scientific fields. This included archeology (Wylie 1989), biology (Blier 1978; Hubbard 1982; Martin 1991), cognitive sciences (Fausto-Sterling 1985), primatology (Hrdy 1986; Schiebinger 2004), and a variety of social sciences (Anderson 2004; Collins 1986; Smith 1979). Feminists argued that there were numerous instances where gendered assumptions and stereotypes influenced observation, description, and interpretation of data. Sarah Hrdy (1986), for example, demonstrated that because of a deeply held evolutionary assumption that males were likely to be promiscuous and females were not, primatologists tended to focus on male primate behavior that confirmed that assumption, while female primate behavior (which was in fact promiscuous) was ignored, downplayed, or interpreted according to expected gender norms. Emily Martin (1991) showed that accounts of human reproduction largely incorporated gendered norms of “active” and “aggressive” sperm and “passive” ovum, and those norms largely obscured researchers from understanding the chemical interactions that occurred between sperm and egg.

Much of this early work was not an attack on the VFI directly but challenged the descriptive claim that science, by and large, operated in accordance with the VFI. The question, for feminists then, was, were these instances isolated cases of “bad science” where political values inappropriately influenced scientific reasoning? Or was there something about the nature of scientific reasoning that might explain why these biases appeared to be widespread and consistent over time, despite the fact that many of the scientists involved were well-trained and well-respected (Code 1991; Haraway 1988; Harding 1986; Longino 1983, 1990; Nelson 1990)? In other words, feminists were interested in examining whether scientific methods and reasoning were objective in the sense previously thought.

Yet despite uncovering ways that values had negatively influenced scientific reasoning in a variety of ways, feminists were less inclined to conclude that biases were caused by values per se. This was for two main reasons. First, feminists were less inclined to think that ethical and political values (e.g., that women should have equal rights and opportunities) were mere subjective preferences or “wishings” and more inclined to think they could be defended on rational grounds (Brown 2020; ChoGlueck and Lloyd 2023; Clough 2003, 2020). Indeed, feminism necessitated believing that some values are more justified than others (e.g., sexist or non-equalitarian ones). Therefore, the problem was not necessarily non-epistemic values as such, but rather that the values that had influenced research were uncritically adopted and, as a result, were often either unjustified or tended to reflect the particular values of the relatively monolithic group of those practicing science (Longino 1990; Rolin 2017).

Second, feminists were less inclined to see the influence of values as isolated cases of “bad science,” because of their other two aims. The fact that greater diversity within science had often led to the identification and evaluation of value-laden assumptions made it seem likely that there was something about the nature of scientific inquiry and knowledge production that was socially situated—or that the social and political values of the context in which science was practiced might be integral to how knowledge is produced. This might require greater diversity in values, approaches, ideas, and experiences among those practicing science, rather than futile attempts of trying to adhere to an unhelpful ideal. That is, feminists were more inclined to consider how values might be managed, rather than avoided, and whether greater representation of underrepresented groups might play an important

role in accomplishing that. Similarly, as feminists began to understand the ways that values operated in science, they wondered whether different values (such as feminist or egalitarian values) could play a positive role and potentially unlock the emancipatory potential of science. This is not to say that they thought feminist values ought to determine the conclusions of science. But they were interested in exploring how feminist values could play beneficial roles.

As a result, feminists were open to considering whether values could be both necessary and desirable in science. In general, feminists launched three kinds of challenges to the VFI. First, feminists argued that factors that influence the so-called context of discovery also have implications for decisions made in the context of justification (Fausto-Sterling 1985; Okruhlik 1994). For example, if researchers are seeking to determine cognitive differences between men and women in order to justify anti-suffrage policies, and such values influence the hypotheses proposed or theories constructed, those theories will determine how concepts are operationalized, judgments about what data is relevant and collected, and what alternative hypotheses or explanations might be considered. That is, values that drive the context of discovery will constrain decision-making related to assessing and evaluating theories (Okruhlik 1994, 38–39). Different values, however, might expand hypotheses, open up new avenues of data or evidence, or contribute to the scope of alternative explanations considered. Similarly, drawing on work from Kuhn (1970) and Quine (1951), feminists argued that evaluating hypotheses or theories required numerous background assumptions that determined whether something consisted of “good evidence” and that non-epistemic value judgments were often some of these background assumptions (Harding 1986; Longino 1990; Nelson 1990). Insofar as these arguments were successful, they suggested that non-epistemic values could not be kept out of theory assessment, and that instead philosophers of science should focus on how to manage such values by, for example, ensuring that such values received critical scrutiny and were themselves justified or by ensuring that there was a diversity of values operating in scientific inquiry. This challenged the practical usefulness of the VFI. If value judgments could not be excluded from the context of justification, pretending that they were not there would likely have negative epistemic consequences because they would not receive critical scrutiny.

A second kind of challenge that feminists made to the VFI was to criticize the purportedly sharp distinction between so-called epistemic values and “non-epistemic” values. Epistemic values were thought to be “epistemic” in virtue of their truth-conduciveness. Yet it is not clear that truth is the only aim of scientific inquiry (Anderson 1995a; Cartwright 1983; Kitcher 2001). Some argued that epistemic values might be expanded to include values such as simplicity or explanatory power that, while not truth conducive, served other cognitive or constitutive aims of science (Kuhn 1977; McMullin 1982). Feminists, however, showed that social and political interests could often be a part of the aims of inquiry (Longino 1996; Rooney 1992). While most feminists agreed that theories had to be empirically adequate, they argued that the selection and weighting of other cognitive values, or which sorts of epistemic tradeoffs are acceptable, can only be judged relative to the interests of the investigation, which might include social or political interests (Longino 1996; Rooney 1992, 2017).

The third challenge feminists raised to the VFI was to question whether scientific inquiry was solely concerned with describing the non-normative natural world. Indeed, many areas of social science, for example, are concerned with social phenomena that may in fact be value-laden. Certain concepts may involve both descriptive and value-laden dimensions. For example, whether something is a disease or a mental illness is not just about certain physical conditions or causal mechanisms but also depends on value judgments about what is bad for human well-being or functioning (Anderson 1995a; Intemann 2001). Several concepts that feminist researchers were particularly interested in investigating or understanding, such as rape, harms from divorce, unemployment, or sexual harassment, all have normative aspects (Anderson 1995a, 2004). Whether to adopt one system of classification over another involves value judgments (that depend on those dimensions we want to understand).

Moreover, background assumptions adopted in investigating these phenomena will likely involve value judgments because the phenomena themselves are value-laden (Intemann 2001).

Even one of these challenges, if successful, would be enough to undermine the VFI (at least the version under consideration here). As Matt Brown has pointed out, the VFI is an “all-or-nothing affair,” and its defense requires that there are no instances of scientific reasoning in the assessment or evaluation of claims that depend on non-epistemic values (Brown 2024, 111).

Yet feminists were interested in challenging more than the VFI. They were interested in challenging the notion of objectivity that motivated the VFI in the first place. Feminists wanted to show that scientific knowledge does, in some sense, reflect the social context (including social and political values) in which it is produced. Moreover, they wanted to show that the social features (including the values) of those who practice science are epistemically significant and that scientific methods can depend on the interests of inquiry (including social, ethical, or political interests). Yet feminists also wanted to deny that this necessarily threatened understanding science as “objective” in some sense. Their aim then became rethinking and reconceptualizing objectivity, and particularly the idea that “objective science” is free of social, ethical, or political values.

2.5 Feminist conceptions of objectivity and alternatives to the VFI

Rather than focusing on alternatives to the VFI, feminists tended to focus on providing alternative accounts of scientific objectivity that could capture the ways in which scientific decision-making could be rational and unbiased without requiring it to be free of social, ethical, or political values. Feminists, of course, also disagreed among themselves about how best to reconceptualize objectivity in light of feminist insights about the roles of values in science.

Helen Longino’s contextual empiricism conceives of scientific objectivity as something that comes in degrees when certain sorts of procedural requirements are met (Longino 1990, 2001). Scientific claims are objectively justified to the extent to which they are subject to a process of transformative criticism. For Longino, this was characterized by there being avenues for criticism, equality of intellectual authority among participants, shared standards of evaluation, and uptake of criticism (Longino 1990). On this account, individual scientists are recognized to have all kinds of personal, political, social, or ethical values and such values may influence them in a variety of ways. Yet at the community level the influence of such values may be scrutinized—both for their content (whether it is shared or reasonable) and for the specific role they are playing.

Other feminists have also offered process-based accounts of objectivity. For example, Heldke and Kellert (1995) argued for thinking of scientific objectivity as responsible inquiry, grounded on work by John Dewey, Donna Haraway, and Lorraine Code. On their account, scientific inquiry is objective to the extent that its participants acknowledge, fulfill, and expand responsibility to the context of inquiry (Heldke and Kellert 1995, 361). Similar to Longino’s view, their account of responsibility involves being responsible to the concerns and criticisms of others, though they place particular importance on responding to those who may be impacted by the research. Again, it does not involve value-freedom but does require meeting epistemic responsibilities.

Sandra Harding’s standpoint account of what she calls strong objectivity is distinct from Longino’s. It focuses more on the cultivation of a critical standpoint, where certain egalitarian values are endorsed, and researchers utilize methodologies aimed at starting with the lives of those disproportionately impacted by research problems (Harding 1992, 1995). Harding argues that research from such standpoints is less partial, because it represents interests that have been historically excluded. As with Longino’s account, it does not require either scientists or scientific reasoning to be free of ethical, social, and political values. Indeed, it centers egalitarian values and emancipatory aims at the heart of the aims of science.

Naomi Scheman (2001, 2011) has developed what might be thought of as a virtue-based account of objectivity as trustworthiness. That is, scientists should strive to be trustworthy in scientific reasoning and the methodologies they use. In particular, they should strive to be trustworthy even for those who might disagree with them. On Scheman's account, this includes being aware and attentive of what different audiences of knowledge claims might care about. That is, it may involve being attentive to their values.

While feminists disagree about how to think about objectivity, what these accounts share is a commitment to understanding objectivity in ways that do not require freedom from political values. Like proponents of the VFI, however, feminists had an interest in protecting science from bias and wanted to maintain that scientific reasoning was rational. This is partly because they wanted to harness the power of science for emancipatory purposes. That is, they recognized that science was an important way to achieve knowledge, but they wanted to improve it so as to address the concerns and needs of those who had been historically underrepresented. As a result, feminists might agree that there should be some constraints on values in science. For example, most feminists acknowledge that political values should never—by themselves—determine the outcome of scientific investigation (e.g., Anderson 1995b). Feminists might also agree that the idiosyncratic values of individual scientists should not be uncritically adopted in scientific reasoning (Longino 1990). Feminists might also agree that values that play a role in science meet some criteria (e.g., they are justified, or widely accepted, or democratically adopted, or benefit those who are marginalized). Of course, in thinking about whether there are alternatives to the VFI, it will be important to consider what function the ideal serves. If it is to protect scientific objectivity, then feminists will be skeptical if it is a conception of objectivity that requires total value-freedom.

2.6 Conclusions

Feminists began challenging the VFI in the 1970s–1990s, arguing that values could play both necessary and important roles in the assessment of scientific theories because (1) values that operate in the context of discovery had implications for decisions in the context of justification, (2) non-epistemic values play a role in selecting, weighing and adjudicating between epistemic or cognitive values for theory choice in a particular research context, and (3) some phenomena that are the object of scientific inquiry are normative in nature. Yet their target of criticism was not only the VFI, but a particular conception of scientific objectivity. The traditional conception of scientific objectivity made feminist experiences and critiques of science unintelligible, because it assumed that objectivity required persons, methods, and the products of science to be value-free and bear no trace of the knower. This conception of objectivity was an obstacle to feminists who wanted to show that (1) science is influenced by the values and social circumstances in which it is practiced, (2) the exclusion of women and other represented groups (and the values they cared about) had negative epistemic consequences, and (3) science might be improved and put to emancipatory purposes with feminist or more egalitarian values. As a result, feminists contributed not only to showing that values were sometimes necessary and important to scientific reasoning. They also contributed to a broader epistemological project of reconceptualizing scientific objectivity such that the VFI was no longer needed or useful.

Notes

- 1 There are potentially other motivations for the VFI, including potentially protecting democracy (see Bright 2018; Douglas and Branch 2024) or facilitating trust with the public (Bright 2018; Odenbaugh 2008). I focus on protecting objectivity here because this is where feminist philosophy of science tended to focus.
- 2 See Douglas (2004, 2009) for a more detailed account of various conceptions of objectivity as related to claims, methods, and persons.

References

- Anderson, Elizabeth. 1995a. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23: 27–58. <https://www.jstor.org/stable/43154207>.
- Anderson, Elizabeth. 1995b. "Feminist Epistemology: An Interpretation and a Defense." *Hypatia* 10 (3): 50–84. <https://doi.org/10.1111/j.1527-2001.1995.tb00737.x>.
- Anderson, Elizabeth. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19 (1): 1–24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>.
- Bleier, Ruth. 1978. "Bias in Biological and Human Sciences: Some Comments." *Signs: Journal of Women in Culture and Society* 4 (1): 159–62. <https://doi.org/10.1086/493577>.
- Bluhm, Robyn. 2020. "Neurosexism and Our Understanding of Sex Differences in the Brain." In *The Routledge Handbook of Feminist Philosophy of Science*, edited by Sharon Crasnow and Kristen Intemann. Routledge.
- Bright, Liam Kofi. 2018. "Du Bois' Democratic Defence of the Value Free Ideal." *Synthese* 195 (5): 2227–45. <https://doi.org/10.1007/s11229-017-1333-z>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Cartwright, Nancy. 1983. *How the Laws of Physics Lie*. Oxford University Press.
- ChoGlueck, Christopher, and Elizabeth A. Lloyd. 2023. "Values as Heuristics: A Contextual Empiricist Account of Assessing Values Scientifically." *Synthese* 201: 220. <https://doi.org/10.1007/s11229-023-04198-z>.
- Clough, Sharyn. 2003. *Beyond Epistemology: A Pragmatist Approach to Feminist Science Studies*. Rowman & Littlefield.
- Clough, Sharyn. 2020. "Using Values as Evidence When There's Evidence for Your Values." *Philosophy in the Contemporary World* 26 (1): 5–37. <https://doi.org/10.5840/pcw2020261/21>.
- Code, Lorraine. 1991. *What Can She Know? Feminist Theory and the Construction of Knowledge*. Cornell University Press.
- Collins, Patricia Hill. 1986. "Learning from the Outsider within: The Sociological Significance of Black Feminist Thought." *Social Problems* 33 (6): s14–s32. <https://doi.org/10.2307/800672>.
- Crasnow, Sharon, and Kristen Intemann. 2024. *Feminist Epistemology and Philosophy of Science: An Introduction*. Routledge.
- Daston, Lorraine, and Peter Gallison. 2007. *Objectivity*. Zone Books.
- Douglas, Heather. 2004. "The Irreducible Complexity of Objectivity." *Synthese* 138: 453–73. <https://doi.org/10.1023/B:SYNT.0000016451.18182.91>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather, and T. Y. Branch. 2024. "The Social Contract for Science and the Value-Free Ideal." *Synthese* 203: 40. <https://doi.org/10.1007/s11229-023-04477-9>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2025. "Characterizing the Value-Free Ideal: From a Dichotomy to a Multiplicity." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw, and Carla Fehr. Boston Studies in the Philosophy and History of Science. Springer.
- Fausto-Sterling, Anne. 1985. *Myths of Gender: Biological Theories about Women and Men*. Basic Books.
- Fine, Cordelia. 2018. "Feminist Science: Who Needs It?" *The Lancet* 392 (10155): 1302–3.
- Fine, Cordelia. 2022. "Fairly Criticized, or Politicized? Conflicts in the Neuroscience of Sex Differences in the Human Brain." *Cold Spring Harbor Perspectives in Biology* 14 (8): a039115. <https://doi.org/10.1101/cshperspect.a039115>.
- Gould, Stephen. J. 1996. *The Mismeasure of Man*. W.W. Norton & Company.
- Gross, Barry. R. 1994. "What Could a 'Feminist' Science Be?" *The Monist* 77 (4): 434–44. <https://www.jstor.org/stable/27903402>.
- Gross, Paul, and Norman Levitt. 1994. *Higher Superstition*. Johns Hopkins University Press.
- Haaek, Susan. 1992. "Science 'From a Feminist Perspective'." *Philosophy* 67 (259): 5–18. <https://doi.org/10.1017/S0031819100039796>.
- Haraway, Donna. 1985. "A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s." *Socialist Review* 80: 65–108.

- Haraway, Donna. 1988. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99.
- Harding, Sandra. 1986. *The Science Question in Feminism*. Cornell University Press.
- Harding, Sandra. 1992. "After the Neutrality Ideal: Science, Politics, and 'Strong Objectivity'." *Social Research* 59 (3): 567–87. <https://www.jstor.org/stable/40970706>.
- Harding, Sandra. 1995. "Strong Objectivity: A Response to the New Objectivity Question." *Synthese* 104: 331–49. <https://doi.org/10.1007/BF01064504>.
- Heldke, Lisa M., and Stephen H. Kellert. 1995. "Objectivity as Responsibility." *Metaphilosophy* 26 (4): 360–78. <https://www.jstor.org/stable/24438975>.
- Hrdy, Sarah B. 1981. *The Woman That Never Evolved*. Harvard University Press.
- Hrdy, Sarah B. 1986. "Empathy, Polyandry, and the Myth of the Coy Female." In *Feminist Approaches to Science*, edited by Ruth Bleier. Pergamon Press.
- Hubbard, Ruth. 1982. "Have Only Men Evolved?" In *Biological Women: The Convenient Myth*, edited by Ruth Hubbard, Mary Sue Henifin, and Fried. Barb. Schenkman Press.
- Intemann, Kristen. 2001. "Science and Values: Are Value Judgments Always Irrelevant to the Justification of Scientific Claims?" *Philosophy of Science* 68 (S3): S506–18. <https://doi.org/10.1086/392932>.
- Jordan-Young, Rebecca M. 2010. *Brainstorm: The Flaws in the Science of Sex Difference*. Cambridge University Press.
- Jorgensen-Earp, Cheryl R., and Darwin D. Jorgensen. 2016. "Physiology and Physical Force: The Effect of Edwardian Science on Women's Suffrage." *Southern Communication Journal* 81 (3): 136–55. <https://doi.org/10.1080/1041794X.2015.1124914>.
- Kitcher, Phillip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kourany, Janet A. 2003. "A Philosophy of Science for the Twenty-First Century." *Philosophy of Science* 70 (1): 1–14. <https://doi.org/10.1086/367864>.
- Kourany, Janet A. 2010. *Philosophy of Science after Feminism*. Oxford University Press.
- Kuhn, Thomas. 1970. *The Structure of Scientific Revolutions*. University of Chicago Press.
- Kuhn, Thomas. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*. The University of Chicago Press.
- Lloyd, Elizabeth. A. 1995. "Objectivity and the Double Standard for Feminist Epistemologies." *Synthese* 104: 351–81. <https://doi.org/10.1007/BF01064505>.
- Longino, Helen. 1983. "Beyond 'Bad Science': Skeptical Reflections on the Value-Freedom of Scientific Inquiry." *Science, Technology, & Human Values* 8 (1): 7–17. <https://doi.org/10.1177/016224398300800103>.
- Longino, Helen. 1987. "Can There Be A Feminist Science?" *Hypatia* 2 (3): 51–64. <https://doi.org/10.1111/j.1527-2001.1987.tb01341.x>.
- Longino, Helen. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Longino, Helen. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn. H. Nelson and Jack Nelson. Springer.
- Longino, Helen. 2001. *The Fate of Knowledge*. Princeton University Press.
- Maney, Donna L. 2016. "Perils and Pitfalls of Reporting Sex Differences." *Philosophical Transactions of the Royal Society B: Biological Sciences* 371 (1688): 20150119. <https://doi.org/10.1098/rstb.2015.0119>.
- Martin, Emily. 1991. "The Egg and The Sperm: How Science Has Constructed a Romance Based on Stereotypical Male-Female Roles." *Signs: Journal of Women in Culture and Society* 16 (3): 485–501. <https://doi.org/10.1086/494680>.
- McMullin, Ernan. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (2): 2–28. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409>.
- Nagel, Ernest. 1979. *The Structure of Science: Problems in the Logic of Scientific Explanation*. Hackett Publishing Company.
- Nelson, Lynn Hankinson. 1990. *Who Knows: From Quine to a Feminist Empiricism*. Temple University Press.
- Odenbaugh, Jay. 2008. "Ecology and the Inescapability of Values." *Science and Engineering Ethics* 14 (4): 593–6. <https://doi.org/10.1007/s11948-008-9092-3>.
- Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24 (sup1): 21–42. <https://doi.org/10.1080/00455091.1994.10717393>.
- Proctor, Robert. 1991. *Value-Free Science?: Purity and Power in Modern Knowledge*. Harvard University Press.
- Quine, Willard V. O. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60 (1): 20–43.
- Reichenbach, Hans. 1938. *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. The University of Chicago Press.

- Richardson, Sarah S. 2013. *Sex Itself: The Search for Male and Female in the Human Genome*. University of Chicago.
- Rolin, Kristina. 2017. "Can Social Diversity Be Best Incorporated into Science by Adopting the Social Value Management Ideal?" In *Current Controversies in Values and Science*, edited by Kevin C. Elliot and Daniel Steel. Routledge.
- Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1): 13–22. <https://doi.org/10.1086/psaprocbsmeetp.1992.1.192740>.
- Rooney, Phyllis. 2017. "The Borderlands between Epistemic and Non-Epistemic Values." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Rossiter, Margaret W. 1982. *Women Scientists in America: Before Affirmative Action 1940-1972 (Volume 2)*. Johns Hopkins University Press.
- Ruse, Michael. 1984. "Biological Science and Feminist Values." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984 (2): 525–42. <https://doi.org/10.1086/psaprocbsmeetp.1984.2.192524>.
- Scheman, Naomi. 2001. "Epistemology Resuscitated: Objectivity as Trustworthiness." In *Engendering Rationalities*, edited by Nancy Tuana and Sandra Morgen. SUNY Press.
- Scheman, Naomi. 2011. *Shifting Ground: Knowledge and Reality, Transgression and Trustworthiness*. Oxford University Press.
- Schickore, Jutta. 2022. "Scientific Discovery." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta & Ur Nodelman. <https://plato.stanford.edu/archives/win2022/entries/scientific-discovery/> (accessed December 12, 2024).
- Schiebinger, Londa. L. 2004. *Nature's Body: Gender in the Making of Modern Science*. Rutgers University Press.
- Smith, Dorothy E. 1979. "A Sociology for Women." In *The Prism of Sex: Essays on the Sociology of Knowledge*, edited by Julia A. Sherman and Evelyn Torton Beck. University of Wisconsin Press.
- Smith, Dorothy E. 1987. *The Everyday World as Problematic: A Feminist Sociology*. Northeastern University Press.
- Valian, Virginia. 1997. *Why So Slow?* MIT Press.
- Wylie, Alison. 1989. "The Interplay of Evidential Constraints and Political Interests: Recent Archaeological Research on Gender." *American Antiquity* 57 (1): 15–35. <https://doi.org/10.2307/2694833>.
- Wylie, Alison, Kathleen Okruhlik, Leslie Thielen-Wilson, and Sandra Morton. 1989. "Feminist Critiques of Science: The Epistemological and Methodological Literature." *Women's Studies International Forum* 12 (3): 379–88. [https://doi.org/10.1016/S0277-5395\(89\)80014-7](https://doi.org/10.1016/S0277-5395(89)80014-7).

3

STANDPOINT THEORY

Kristina Rolin

3.1 Introduction

This chapter discusses standpoint theory as a naturalized approach to philosophy of science and focuses on *feminist* standpoint theory as an example. Naturalized philosophy of science aims to understand which means are conducive to achieving desired epistemic goals (Giere 1989). The most important, even though not the only, epistemic goal for standpoint theory is to understand how marginalization and oppression are maintained in society and how they can be resisted and unraveled. In the spirit of naturalism, standpoint theory asks which social conditions are conducive to knowledge about the workings of marginalization or oppression in society. According to standpoint theory, the social location of researchers as well as research participants is an epistemic resource which can generate novel evidence, interpretations of evidence, theoretical concepts, criticism of existing research, and scientific breakthroughs. As Wylie argues, standpoint theory is not meant to be a “full-service epistemology” (2012, 61). It is designed to serve the purpose of understanding how socially relevant and potentially emancipatory knowledge can be produced under conditions of unjust social inequalities.

Like other naturalized approaches to philosophy of science, standpoint theory questions the assumption that the value-free ideal of science is necessary for achieving desired epistemic goals, especially if science is to achieve these goals in responsible ways. According to standpoint theory, the value-free ideal—the view that non-epistemic values are not allowed to play a role in the acceptance of hypotheses and scientific theories—is not feasible. This is because all scientific knowledge is socially situated in ways that are likely to imply tacit value judgments concerning, for example, which concepts are used, how they are defined, and what kind of evidence is relevant to the inquiry. Standpoint theory is also committed to the view that the value-free ideal is not desirable because insofar as value judgments are inevitable in the core activities of scientific inquiry, they should be based on the “right” values. According to standpoint theory, it is not enough to identify and criticize problematic value influences in scientific research. It is equally important to ensure that scientific research is guided by morally and socially justifiable values. Thus, standpoint theory provides not only an account of the proper role of moral and social values in science but also an account of what the proper values are (e.g., anti-racism, anti-sexism, and other egalitarian values).

While standpoint theory is consistent with other criticisms of the value-free ideal, it departs from them by emphasizing the role of moral and social values in the production of evidence. For example,

standpoint theory differs from the inductive risk argument by calling attention to what happens before evidence is in rather than after it is in. The inductive risk argument claims that moral and social values are necessary to responsible value judgments of acceptable uncertainty in the acceptance of hypotheses (see [Elliott 2026](#)). Whereas the inductive risk argument focuses on the role of moral and social values in evidential reasoning that takes place after the evidence is in, standpoint theory suggests that moral and social values should also guide efforts to generate evidence under social conditions that are likely to impede scientific inquiry.

Standpoint theory thus poses a challenge to the value-free ideal by suggesting that what happens before evidence is in is as important to the acceptance of hypotheses and scientific theories as what happens after evidence is in. This is because what kind of evidence is collected and taken seriously will eventually influence the assessment of the evidential support for hypotheses and scientific theories (see also [Elliott and McKaughan 2009](#)). From the point of view of naturalized philosophy of science, it is question-begging to assume that the collection of evidence is part of the “context of discovery” rather than the “context of justification” especially in light of the view that the collection of evidence is as relevant to the epistemic success of inquiry as reasoning from evidence to hypotheses and scientific theories. As standpoint theory questions a clear-cut distinction between the “context of discovery” and the “context of justification,” it can be seen as a version of the “blurred boundaries” argument against the value-free ideal (see [Elliott 2026](#)).

Standpoint theory also differs from critical contextual empiricism by emphasizing the necessity of including the perspectives of marginalized social groups into scientific communities. Critical contextual empiricism claims that moral and social values can legitimately play a role in judgments concerning what kind of evidence is relevant to a hypothesis or a theory (see [Borgerson 2026](#)). As [Longino \(1990\)](#) argues, moral and social values are allowed to influence the choice of background assumptions that are needed to establish the relevance of empirical evidence to a hypothesis or a theory as long as no one has refuted these assumptions. Critical contextual empiricism calls for a diversity of perspectives because such diversity is likely to engender critical exchanges in scientific communities and help researchers detect value-laden background assumptions. While standpoint theory also values criticism, it takes one step further by suggesting that the perspectives of marginalized social groups are necessary to challenge dominant perspectives which tend to serve the interests of privileged social groups. To challenge dominant perspectives, it is not sufficient to criticize value-laden background assumptions; it is necessary to gather new evidence and foster new perspectives.

This chapter aims to clarify how standpoint theory, and especially feminist standpoint theory, perceives the proper role of moral and social values in scientific inquiry. The chapter is divided into three sections. Each section discusses one of the three main theses of standpoint theory: The situated knowledge thesis, the thesis of epistemic advantage, and the achievement thesis. The conclusion summarizes the discussion.

3.2 The situated knowledge thesis

According to the situated knowledge thesis, all scientific knowledge is socially situated ([Harding 1991](#), 11; [2004](#), 7). As Wylie argues, “social location systematically shapes and limits what we know, including tacit, experiential knowledge as well as explicit understanding, what we take knowledge to be as well as specific epistemic content” ([2003](#), 31). She also argues that it is a matter of empirical inquiry to find out how social location shapes our experience in a particular context and how this experience is relevant to specific research projects (32). The situated knowledge thesis does not assume that all members of a social group have similar life experiences. It can accommodate differences among group members by suggesting that on average they encounter similar situations. Thus,

membership in certain social groups is often relevant to understanding the kinds of experience we have because it affects our social, political, and material circumstances (Intemann 2010, 785).

Wylie (2012) introduces a distinction between generic and systemic situated knowledge theses. The generic thesis is the view that there is no such thing as a “view from nowhere” since contingent histories and social contexts affect what one can know and what is recognized as knowledge (Wylie 2012, 61). The systemic thesis is the view that the production of scientific knowledge is both enabled and constrained by systemic relations of power that prevail in society at large (Wylie 2012, 62). Systemic relations of power mean that members of one social group, the dominant group, can limit the choices available to members of another social group, the subordinate group. Systemic relations of power are vehicles of domination when they constrain choices in ways that are harmful for members of the subordinate group. While a version of the generic situated knowledge thesis can be found in many naturalized and social approaches to philosophy of science, it is the systemic thesis that distinguishes standpoint theory from other approaches.

Given the systemic situated knowledge thesis, standpoint theory is a social epistemology of scientific knowledge that focuses on the study of power relations and how they affect the production of scientific knowledge (Rolin 2009). Systemic relations of power pose special challenges to social scientific inquiry because they can suppress or distort relevant evidence (Rolin 2009, 219). As they can be used to dominate people, they are likely to mobilize a complex set of motivations that prompt people to either conceal or distort relevant evidence (Rolin 2009). Consequently, not just any social location is of interest to standpoint theory but those social locations that track systemic relations of power. Gender, race, and class have received much attention because in many national and cultural contexts, economic benefits and other types of privileges are distributed along lines marked by these social categories. In standpoint theory, the situatedness of knowers is understood in structural terms rather than in terms of individual or idiosyncratic ways of seeing the world.

In the context of scientific research, the social location of the researcher as a member of a historically underrepresented social group in academia can be epistemically significant in many ways. In an influential article “Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought,” Collins (1986) argues that the outsider-within status enables Black feminist intellectuals to combine their first-hand experience of marginalized social locations with their insider perspective on academic knowledge production. They can also draw on the longer history of the outsider within status as numerous Black women were insiders of white families in their role as domestic workers, while remaining outsiders in virtue of their race. The social location of the outsider within enables one to develop a critical posture toward research that either ignores or distorts the reality of marginalized social groups. For example, sociological research can challenge externally defined, controlling images of Black women in the U.S. (Collins 1986, S17).

It is not only the social location of the researcher that is of epistemic interest. The systemic situated knowledge thesis is also relevant when researchers recruit research participants or build research collaborations. Studying marginalized social groups can provide valuable insights into the mechanisms of marginalization and oppression, but it also raises challenging questions. For example, researchers should reflect on how systemic relations of power shape interactions between research participants and themselves as privileged academics. They also need to decide whether they invite research participants to participate in the design of the study, the interpretation of the data, and other stages of research beyond merely agreeing to be interviewed or observed. Researchers should also consider how they can carry out the study without unduly burdening research participants who are already struggling in their daily lives (Rolin 2025).

The systemic situated knowledge thesis is of interest to values and science debates in two ways. First, the thesis suggests that scientific knowledge is likely to be value-laden because researchers’ social locations can prevent them from recognizing how values shape their research questions, central

concepts, theoretical frameworks, background assumptions, or evidential reasoning. Second, the thesis suggests that researchers' social location is an epistemic resource that helps them detect value-laden assumptions especially when the assumptions are at odds with their experiences. Like Longino (1990), standpoint theorists argue that scientific communities should include diverse perspectives to raise awareness of value influences in science and to generate critical debates about them.

As standpoint theory is concerned with systemic relations of power and their impact on the production of knowledge, it recommends that researchers gather evidence in marginalized social locations and with people inhabiting these locations to understand the causes and consequences of marginalization and oppression in society. Marginalized social locations can be a source of insight into the way power relations work in society as well as in academia (Wylie 2012, 63). The next section discusses the epistemic benefits of starting an inquiry in marginalized social locations.

3.3 The thesis of epistemic advantage

The thesis of epistemic advantage claims that marginalized social locations give rise to epistemic benefits in particular research projects (Wylie 2003, 26). I have argued that the thesis of epistemic advantage is to be understood as an empirical hypothesis (Rolin 2006). Case examples can be used to examine what the epistemic benefits are and how they are related to the experience of those occupying marginalized social locations. For example, feminist research on gender-based discrimination in academia suggests that an epistemic advantage can be a conceptual innovation (e.g., the notion of subtle forms of discrimination) which draws attention to new bodies of evidence ignored in earlier research on women's underrepresentation in science (Rolin 2006, 130).

Wylie argues that the thesis of epistemic advantage should not be understood as "an attribution of automatic or comprehensive epistemic privilege to members of subdominant social groups" (2012, 62). Insofar as members of marginalized social groups have an epistemic advantage when it comes to understanding the mechanisms of marginalization and oppression, the advantage is in a contingent relation to their social location. Membership in a marginalized social group is not sufficient for having a less distorted view of social reality. Also, a standpoint developed in a particular social location is unlikely to be relevant to every scientific field or every social and cultural context. Yet, standpoint theorists do not see this as a weakness of standpoint theory because they believe that all social scientific research provides a local and partial perspective on social reality even when it pretends to be universal and impartial. In this respect, standpoint theory shares Haraway's criticism of the myth-like understanding of objectivity as "the god-trick of seeing everything from nowhere" (1991, 189).

When the thesis of epistemic advantage is understood as an empirical hypothesis, it raises the question of how marginalized social locations come to have epistemically advantageous consequences, if not always, at least under some circumstances, and for some research projects. Whereas many philosophers of science have turned to case examples to analyze epistemic advantages emerging from marginalized social locations (Rolin 2006; Wylie 2003), standpoint theory has also proposed several hypotheses to explain the connection between marginalized social locations and epistemic advantages. According to one hypothesis, epistemic advantage arises not only because marginalized people on average have personal experiences of oppression but also because they have a stronger interest and motivation to understand the workings of oppression than privileged people have. To navigate their way in oppressive social circumstances, marginalized people must understand society from the dominant group's perspective, whereas members of the dominant group can and often do ignore the perspectives of the oppressed.

Another hypothesis suggests that the epistemic advantage is due to testimonial injustice suffered by members of marginalized social groups. Wu argues that insofar as their testimonies are ignored or devalued by the dominant group, their perspectives are sources of genuinely novel insights

(2023, 756). Thus, marginalized social locations can help one discover gaps in existing scientific knowledge and initiate novel lines of inquiry to remedy epistemic injustices.

Yet another hypothesis suggests that marginalized social locations give rise to epistemic advantages because they provide access to shared experiences of marginalization. As [Dror \(2023\)](#) argues,

Insofar as members of marginalized groups tend to live together, trust each other more than they trust members of dominant groups, and thus are more comfortable discussing their experiences with each other, marginalized people will also be better placed to hear the experiences of other similarly marginalized people. (622)

Dror also argues that the empirical and thus contingent nature of the thesis of epistemic advantage has an important implication for privileged social groups. They are not epistemically disadvantaged in principle and are thus often blameworthy for their ignorance about oppression ([Dror 2023](#), 618).

In sum, standpoint theory holds the view that moral and social values should guide research questions and research design from the very beginning of the inquiry so that researchers give due attention to the perspectives of marginalized social groups. The epistemic advantage of marginalized social locations lies in their capacity to generate evidence relevant to understanding systemic relations of power. It goes without saying that standpoint theory does *not* suggest that marginalization is to be maintained because it can generate epistemic benefits. Marginalized social locations are epistemically valuable not so much for the sake of knowledge itself but rather because the kind of knowledge they generate can serve social justice and end marginalization.

3.4 The achievement thesis

The achievement thesis is the view that insofar as there is an epistemic advantage associated with marginalized social locations, the advantage is achieved from a standpoint ([Crasnow 2014](#), 157–60; [Harding 2004](#), 8; [Intemann 2010](#), 785; [Wylie 2003](#), 31). Occupying a particular social location or viewing the world from a particular perspective does not yet amount to having a standpoint ([Harding 1991](#), 123). Feminist standpoint theory emphasizes the role of critical awareness, commitment to egalitarian values, and community building in bringing about the epistemic benefits that marginalized social locations can offer. Of the three main theses of standpoint theory, the achievement thesis is the most controversial. For example, Bright argues that “In so far as one stresses the need for achieving something beyond mere membership in the group, one divorces the standpoint more and more from what people in the pertinent group actually think” (2024, 85).

According to standpoint theory, a standpoint differs from a social location or a mere perspective in three ways. First, to achieve a standpoint one must be critically aware of the social conditions under which knowledge is produced and how these conditions both enable and constrain the production of knowledge ([Intemann 2010](#), 785; [Wylie 2003](#), 31; [2012](#), 63). Second, the knowledge produced from a standpoint must serve the interests of marginalized social groups and aim for social justice ([Crasnow 2014](#), 156; [Harding 1991](#), 126–27; [Intemann 2010](#), 786). Third, a standpoint is a collective rather than an individual achievement ([Intemann 2010](#), 786; [Wylie 2012](#), 49).

In feminist standpoint theory, the achievement thesis is often understood to mean that feminist values should inform not only the aims of scientific research but also the kind of activism that researchers are engaged in ([Intemann 2010](#), 786). Feminist consciousness-raising groups provide an example of how groups of women were able to reflect on their individual experiences and identify patterns they were not able to recognize on their own (786). Many conceptual resources of feminism, such as the notion of sexual harassment, can be attributed to consciousness-raising groups ([Toole 2021](#), 342). The achievement of a standpoint is like the development of expertise in a particular domain

(Saint-Croix 2020; Toole 2024). Thus, a standpoint can be seen as a kind of experience-based expertise that lay people can have.

In the context of scientific research, achieving a standpoint involves the project of building epistemic communities (Rolin 2006, 135). Crasnow argues that standpoints are achieved by epistemic communities which are also “political” communities in the sense that they are built on shared interests (2014, 159). Shared interests should not be taken for granted but understood as an outcome of negotiation and coalition building (Crasnow 2014, 159). The achievement thesis suggests that insofar as there is an epistemic advantage associated with marginalized social locations, the advantage depends on a value- and interest-based epistemic community.

I have developed the achievement thesis further by arguing that feminist standpoint theory can be understood as a social epistemology of scientific/intellectual movements (Rolin 2016). Scientific/intellectual movements (in brief, SIMs) are “collective efforts to pursue research programs or projects for thought in the face of resistance from others in the scientific or intellectual community” (Frickel and Gross 2005, 206). As Kuhn (1996) has proposed, scientific change should be understood not merely as a cognitive change but also as a social phenomenon whose dynamics are in important ways like social movements. But unlike Kuhn, standpoint theorists suggest that the social dynamics of scientific change involve systemic relations of power and collective attempts to challenge them. Moral and social values play an important role in motivating and mobilizing researchers to participate in scientific/intellectual movements. I have argued that “The *feminism* in feminist standpoint theory is not merely about doing scientific research with an eye on egalitarian and emancipatory social reforms; it is also about building SIMs” (Rolin 2016, 12).

A feminist scientific/intellectual movement can be epistemically significant in two ways. First, it enables researchers to generate evidence under conditions where systemic relations of power tend to suppress or distort evidence. This is called “evidence-by-empowerment thesis” because it suggests that the process of gathering evidence should be coupled with the process of empowerment to produce novel evidence despite suppressive social conditions (Rolin 2016, 16–17). Second, a feminist scientific/intellectual movement provides researchers with an epistemic community where they can receive fruitful criticism for research which may be ignored by larger scientific community. This is called “community-by-empowerment thesis” because it suggests that hostile or indifferent epistemic communities are best resisted by forming new epistemic communities where researchers can give and receive constructive criticism (Rolin 2016, 17). Empowerment here is understood as one’s ability to act despite or in response to the power wielded over one by others. While an individual researcher can work on justice-oriented research projects, there are limits to what she can achieve on her own. A scientific/intellectual movement has the benefit that it can challenge systemic relations of power within academia as well as in society, thereby exceeding the powers of individuals.

3.5 Conclusion

In sum, standpoint theory holds the view that the value-free ideal of science is neither feasible nor desirable. It is not feasible because socially situated knowledge is likely to be value-laden in many ways. For example, moral and social values can play a role in which concepts are employed, and which background assumptions are used to support reasoning from evidence to hypotheses and scientific theories. As these choices are relevant to the assessment of the acceptability of hypotheses and scientific theories, it can be difficult to achieve the value-free ideal.

Standpoint theory also holds the view that the value-free ideal is not *morally* desirable because responsible researchers should make morally and socially justifiable value judgments rather than pursue the elusive goal of value freedom. According to standpoint theory, scientific research should serve morally and socially valuable goals, such as equality, non-discrimination, and social justice.

Standpoint theorists believe that for research to serve these goals, it should aim for empirically adequate theories (Wylie 2003).

Standpoint theory is also committed to the view that the value-free ideal is not *epistemically* desirable because the production of evidence should be guided by moral and social values to ensure that evidence is not suppressed or distorted by systemic relations of power. Relevant evidence about the mechanisms of marginalization and oppression does not just lie out there in the world waiting for researchers to come by and document it (Rolin 2016, 18). In many cases, researchers need to challenge systemic relations of power to be able to gather such evidence. This may mean, for example, challenging established practices for collecting statistical data. Or it may mean that researchers build relations of trust with members of marginalized social communities to support their research participation. Thus, standpoint theory strengthens the case against the value-free ideal of science by explaining how moral and social values can play an epistemically productive role in scientific inquiry.

While feminist standpoint theory is often seen as a competing approach to feminist empiricism (Harding 1986), current versions of feminist standpoint theory and feminist empiricism, such as critical contextual empiricism, have much in common (Intemann 2010). Both approaches aim to identify and minimize gender, race, and class bias in scientific research, explain why the inclusion of historically underrepresented social groups is epistemically beneficial for science, and establish a positive role for feminist values in research. Intemann argues that contemporary feminist standpoint theory is properly understood as a form of empiricism, *feminist standpoint empiricism*, because it involves the view that experience is central in the justification of scientific knowledge (Intemann 2010, 785). Standpoint theory and feminist empiricism are both versions of social epistemology because they emphasize the role of epistemic communities in the justification of scientific knowledge (Intemann 2010, 787). Both views are also committed to a kind of contextualism, the view that the justification of scientific knowledge takes place in a particular context of background assumptions and values that guide evidence-seeking and the interpretation of evidence (Intemann 2010, 787).

The empiricism of standpoint theory is a key to understanding why social locations are of epistemic interest and why some social locations offer better insights than some others. Social locations matter because they provide access to certain kinds of empirical evidence. Being marginalized in society means that one is likely to have experiences that people in dominant social groups do not have. When these experiences become evidence that is relevant to scientific theorizing about society, they can change both theory and practice.

References

- Borgerson, Kirstin. 2026. "Critical Contextual Empiricism." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Bright, Liam Kofi. 2024. "Duboisian Leadership through Standpoint Epistemology." *The Monist* 107 (1): 82–97. <https://doi.org/10.1093/monist/onad032>.
- Collins, Patricia Hill. 1986. "Learning from the Outsider within: The Sociological Significance of Black Feminist Thought." *Social Problems* 33 (6): 14–32. <https://doi.org/10.2307/800672>.
- Crasnow, Sharon. 2014. "Feminist Standpoint Theory." In *Philosophy of Social Science: A New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi. Oxford University Press.
- Dror, Lidal. 2023. "Is There an Epistemic Advantage to Being Oppressed?" *Noûs* 57 (3): 618–40. <https://doi.org/10.1111/nous.12424>.
- Elliott, Kevin C. 2026. "Arguments against the Value-Free Ideal." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. "How Values in Scientific Discovery and Pursuit Alter Theory Appraisal." *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Frickel, Scott, and Neil Gross. 2005. "A General Theory of Scientific/Intellectual Movements." *American Sociological Review* 70 (2): 204–32. <https://doi.org/10.1177/000312240507000202>.

- Giere, Ronald N. 1989. "Scientific Rationality as Instrumental Rationality." *Studies in History and Philosophy of Science Part A* 20 (3): 377–84. [https://doi.org/10.1016/0039-3681\(89\)90013-7](https://doi.org/10.1016/0039-3681(89)90013-7).
- Haraway, Donna. 1991. "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective." In *Simians, Cyborgs, and Women: The Reinvention of Nature*. Free Association Books.
- Harding, Sandra. 1986. *The Science Question in Feminism*. Cornell University Press.
- Harding, Sandra. 1991. *Whose Science? Whose Knowledge? Thinking from Women's Lives*. Cornell University Press.
- Harding, Sandra. 2004. "Introduction: Standpoint Theory as a Site of Political, Philosophical, and Scientific Debate." In *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies*, edited by Sandra Harding. Routledge.
- Intemann, Kristen. 2010. "25 Years of Feminist Empiricism and Standpoint Theory: Where Are We Now?" *Hypatia* 25 (4): 778–96. <https://doi.org/10.1111/j.1527-2001.2010.01138.x>
- Kuhn, Thomas. 1996. *The Structure of Scientific Revolutions*. 3rd ed. The University of Chicago Press.
- Longino, Helen. 1990. *Science as Social Knowledge*. Princeton University Press.
- Rolin, Kristina. 2006. "The Bias Paradox in Feminist Standpoint Epistemology." *Episteme* 3 (1–2): 125–36. <https://doi.org/10.3366/epi.2006.3.1-2.125>.
- Rolin, Kristina. 2009. "Standpoint Theory as a Methodology for the Study of Power Relations." *Hypatia* 24 (4): 218–26. <https://doi.org/10.1111/j.1527-2001.2009.01070.x>
- Rolin, Kristina. 2016. "Values, Standpoints, and Scientific/Intellectual Movements." *Studies in History and Philosophy of Science Part A* 56: 11–9. <https://doi.org/10.1016/j.shpsa.2015.10.008>.
- Rolin, Kristina. 2025. "Resisting Epistemic Exploitation: From Institutional Remedies to Community-Based Research." *Studies in History and Philosophy of Science* 111: 63–71. <https://doi.org/10.1016/j.shpsa.2025.05.005>.
- Saint-Croix, Catharine. 2020. "Privilege and Position." *Res Philosophica* 97 (4): 489–524. <https://doi.org/10.11612/resphil.1953>.
- Toole, Briana. 2021. "Recent Work in Standpoint Epistemology." *Analysis* 81 (2): 338–50. <https://doi.org/10.1093/analys/anab026>.
- Toole, Briana. 2024. "Standpoint Epistemology and Epistemic Peerhood: A Defense of Epistemic Privilege." *Journal of the American Philosophical Association* 10 (3): 409–26. <https://doi.org/10.1017/apa.2023.6>.
- Wu, Jingyi. 2023. "Epistemic Advantage on the Margin: A Network Standpoint Epistemology." *Philosophy and Phenomenological Research* 106 (3): 755–77. <https://doi.org/10.1111/phpr.12895>.
- Wylie, Alison. 2003. "Why Standpoint Matters." In *Science and Other Cultures: Issues in Philosophies of Science and Technology*, edited by Robert Figueroa and Sandra Harding. Routledge.
- Wylie, Alison. 2012. "Feminist Philosophy of Science: Standpoint Matters." *Proceedings and Addresses of the American Philosophical Association* 86 (2): 47–76. <https://www.jstor.org/stable/43661298>.

4

VALUES AND OBJECTIVITY

Inkeri Koskinen

4.1 Introduction

In contemporary societies, objectivity is a contested concept used in many fields. Objectivity is often demanded not only from science but also from the judiciary, from decision-making authorities, and from journalists. This demand is typically both epistemic and moral—Lorraine Daston and Peter Galison (2007, 39) use the apt expression “moralized epistemology” when describing the emergence of objectivity as a central virtue in science in the nineteenth century. But the demand for objectivity is also often understood as a requirement to disregard human values: According to this view, objective research would be value-free, and objective decision-making would not be influenced by an individual’s desires or preferences. The demand for objectivity thus often implies that it is one’s moral duty to set aside one’s values. Unsurprisingly, this vaguely self-contradictory demand has sparked much suspicion and criticism: According to the critics, it is neither possible nor worth striving for as an ideal.

In this chapter, I primarily focus on how the uneasy relationship between objectivity and values has been addressed in philosophy of science in the recent decades. Objectivity is discussed and the concept is also used in other areas of philosophy, and sometimes these discussions and uses differ from how the notion of objectivity is understood and used in philosophy of science. I will briefly return to some of these differences at the end of the chapter. As to the question of what exactly we mean by “values,” I will leave it to the other authors of this *Handbook*; here it suffices to say that philosophers working on the topic have repeatedly noted that in the literature on values in science the word “value” is used in alarmingly many ways (Biddle 2013; Elliott 2025; Hilligardt 2022; Solomon 2012; Ward 2021).

I will begin by introducing two accounts of objectivity that have been central to discussions of objectivity in the philosophy of science in recent decades. I will then move on to the numerous arguments that have led to the nearly unanimous rejection of these two accounts.

4.2 Objectivity as faithfulness to facts and objectivity as value-freedom

The philosophical discussion of objectivity in science has, over the last half-century, revolved around criticisms of two influential accounts of objectivity: Objectivity as faithfulness to facts,¹ and objectivity as value-freedom. Many of these criticisms include the claim that non-epistemic values play important roles in scientific research. While recently the target of the criticisms has mostly been the

idea of objectivity as value-freedom, many of the important criticisms challenge both accounts. This is the case particularly in the postcolonial and feminist criticisms that predate the current discussions about values in science. I will now summarize the two criticized accounts and then outline arguments that have been presented against them which highlight the role of values in science in some way or another.

The first of these two oft-criticized accounts identifies objectivity as faithfulness to facts. Bernard Williams (1985) has named this the “absolute conception”: Objective knowledge claims present the world as it is, free from distortions caused by human subjectivity. This conception of objectivity thus presupposes that there is a world that is independent of us, and that science, when successful, produces true knowledge about this world. It links objectivity to representation and to truth in the sense of correspondence: An objective representation of the world is true because it corresponds to the way things are.

The idea of objectivity as faithfulness to facts often includes the idea of aperspectivity: Objective knowledge about, say, the height of a tree does not depend on the direction from which we see it. In the words of Thomas Nagel (1986), this conception of objectivity presupposes a “view from nowhere.” Although few philosophers of science have defended the idea of objectivity as faithfulness to facts (for a recent counterexample, see Hoyningen-Huene 2023), the requirement of perspective-independence has influenced some discussions, for example, about whether the human sciences can produce objective knowledge. Since both the existence and the characteristics of the phenomena that human sciences study are dependent on humans, how could human sciences produce knowledge that conforms to the absolute conception of objectivity (see Montuschi 2003)?

The other influential but contested account of objectivity claims that objective research is value-free. One of the influential early articulations of this understanding of objectivity is Max Weber’s (1904) idea that although values are bound to guide the choice of the questions asked by the social sciences, the research itself must be value-free. In its current form, however, the value-free ideal is relatively recent. Douglas (2009, chap. 3) argues that it only emerged in the Cold War period, and that its function was to argue that science is essentially apolitical. One of its clearest advocates in that period, Isaac Levi, argued that scientists commit themselves to scientific “canons of inference” or “epistemic values,” and given a scientist’s commitment to these, “he need make no further value judgments in order to decide which hypotheses to accept and which to reject” (Levi 1960, 356).

This widely recognized and widely questioned formulation of the value-free ideal demands that non-epistemic values must not influence the key internal stages of objective research: The collection of evidence and the acceptance or rejection of theories or hypotheses. There are two reasons for this adjustment. First, it would be absurd to demand that values should not influence the choice of research topics and questions or the use of the results. Second, following Thomas Kuhn (1977), philosophers generally accept the claim that there are also epistemic values. For example, the fruitfulness, predictive power, accuracy, simplicity, and internal coherence of a theory are values, and scientists often have to choose which ones of them to maximize: An accurate theory is not necessarily a simple one. No proponent of the value-free ideal has ever claimed that this would threaten the objectivity of science. Therefore, the fairest way to express the idea that objective research is value-free is to say that according to it, only epistemic values are allowed to influence the internal stages of a research process. Non-epistemic values—such as moral, social, or political values—may only influence the selection of research topics, the formulation of research questions, and the use of the results (Brown 2024; Douglas 2009).

To understand how these two accounts are connected it is useful to look briefly into the (surprisingly brief) history of the modern notion of objectivity. As Lorraine Daston and Peter Galison (2007) have shown, it is post-Kantian. Although the words “objectivity” and “objective” had been in use for centuries, they were not common and had very different meanings from those

familiar to us today: In scholastic usage, “objective” referred to “things as they are presented to consciousness” (Daston and Galison 2007, 29). Objectivity emerged as a virtue in science only in the nineteenth century. This happened because the understanding of human agency, or the self, had changed: People were no longer seen as passive recipients of impressions but as active agents when observing the world. Daston and Galison (2007, 34) describe how this post-Kantian idea raised a concern: The willful, active, subjective self of the scientist might “prettify, idealize, and, in the worst case, regularize observations to fit theoretical expectations.” In science, it was centrally the distortion of representations that was feared. This gives us the first of the two accounts summarized above, objectivity as faithfulness to facts: An objective representation is not distorted by human subjectivity. The second account focuses on the objectivity of research processes rather than on the objectivity of representations and identifies values as the source of the threatening distortions.

4.3 The intertwined criticisms

These two accounts of objectivity—the idea of objectivity as faithfulness to facts and the idea of objectivity as value-freedom—have been criticized both together and separately. Not all of these criticisms are related to values. For instance, ideas about the empirical underdetermination of scientific theories and the theory-ladenness of observations, and the recognition of scientific terms having clear meanings only within some theoretical framework, have led to the unpopularity of those accounts of objectivity that imply there to be aperspectival knowledge. These arguments do not address questions about values in science. However, the idea of aperspectivity can be connected to the idea of value-freedom: If having a perspective is understood as having a value-laden perspective, then criticisms of one of the influential accounts can also be understood as criticisms of the other. As it remains somewhat unclear what it means to be “aperspectival,” and as we do not even have enough clarity about what we mean by “values,” it is sometimes difficult to determine whether some criticism of the idea of objectivity as faithfulness to facts should be read as a criticism of idea of objectivity as value-freedom, even if it does not explicitly address questions about the appropriate role of values in science. In this section, however, I focus on criticisms that target one or both of the influential accounts and directly address issues related to values. I will start by introducing an argument which postcolonial and feminist critics have formulated in numerous versions and then continue to the arguments against the two accounts presented in philosophy of science.

Postcolonial and feminist critics have noted that the claimed objectivity of a scientific method, a court ruling, or a journalistic report often systematically privileges the interests of the powerful. As Frantz Fanon (1961/2004) put it, “[f]or the colonized subject, objectivity is always directed against him” (37). Along the same lines, Edward Said (1978) noted that the claimed objectivity of Western historians was sheer political propaganda, “the implication always being that Muslims and Arabs cannot be objective but that Orientalists like Lewis writing about Muslims and Arabs are, by definition, by training, by the mere fact of their Westernness” (319). Some feminists have similarly contested the alleged virtue of objectivity. Catharine A. MacKinnon (1987) summarized the gist of these protests by saying that “[o]bjectivity is the epistemological stance of which objectification is the social process, of which male dominance is the politics, the acted-out social practice” (50).

These complaints have in common the claim that the notion of objectivity serves in reality to bolster the position of those in power, and that this is the actual social function of using the concept. Some of the critics—for instance, MacKinnon—argue that this is all there is to objectivity, and that the ideal should be rejected. Others want to demonstrate that what is claimed to be objective is in fact often anything but objective. The argument often targets both the idea of objectivity as faithfulness to facts and the idea of objectivity as value-freedom, since both of these mean that a claim of

objectivity is a claim of neutral aperspectivity, which helps to mask the unjust exercise of power (see [Hawkesworth 1994](#); [Toole 2022](#)).

In feminist philosophy of science, the notion of “the god trick of seeing everything from nowhere” ([Haraway 1988](#), 581) has received extensive criticism. It often incorporates some elements of the complaints sketched above. Elizabeth [Lloyd \(1995\)](#), for instance, has noted that many philosophers have inconsistently blamed feminist philosophy for a lack of objectivity. As she noted, in contemporary secular philosophy the idea of objectivity as aperspectival, impartial knowledge of the “Really Real” has been largely abandoned as untenable. But when criticizing feminist philosophy, many philosophers still implicitly appealed to these ideals. In such cases the concept is clearly used as a kind of a rhetorical weapon in the unjustified marginalization of critical voices.

Standpoint epistemology’s emphasis on the epistemic significance of the standpoints of socially marginalized people and groups challenges the ideal of aperspectivity in a particularly clear manner. One of the cornerstones of standpoint theory is the situated knowledge thesis: Our position in the world strongly influences what we come to know. Socially marginalized groups may collectively develop knowledge and understandings of, for instance, mechanisms that perpetuate inequality—mechanisms that are practically invisible to those in more privileged positions. Recognizing such knowledge and lending support to its development can help researchers identify and understand insufficiently studied questions and phenomena. To succeed in this, it is not enough to admit the potential epistemic value of the perspectives of socially marginalized groups. A feminist standpoint theorist should get involved in the “creation of groups’ consciousness” ([Harding 2004a](#), 32) by participating in the development of such perspectives into epistemically valuable standpoints. The aim is to amend a situation where academic research is systematically skewed toward questions that seem important from the perspective of those in power. Knowledge claimed to be aperspectival is once again likely to be produced from a position of power, and thus likely to bolster existing inequalities ([Crasnow 2013](#); [Intemann 2010](#); [Rolin 2009](#); [Wylie 2003](#)).

As noted, in the recent decades, discussions about objectivity in philosophy of science have mostly revolved around several criticisms of the idea of objectivity as value-freedom. It has been criticized for being conceptually unclear, impossible to follow, and flawed as an ideal: Scientists must allow non-epistemic values to influence all stages of their work.

First, the idea of objectivity as value-freedom presupposes a clear distinction between epistemic and non-epistemic values. When [Kuhn \(1977\)](#) introduced the concept of epistemic values, he noted that scientists often have to choose which epistemic values to prioritize in their work. Helen [Longino \(1996](#); see also [Rooney 1992](#)) has pointed out that this choice is not necessarily made on epistemic grounds. For example, a preference for simplicity may stem from a researcher’s inclination toward theories that present the world as less complex, and this preference might be influenced by the researcher’s political values. If so, should we take simplicity to be an epistemic or a political value?

[Longino \(1990, 2001\)](#) also argues that the aim of value-freedom is unachievable. Science simply cannot be value-free in the way the ideal suggests. Longino bases her argument on a Kuhnian argument for underdetermination and on the observation that scientific research always rests on a large number of background assumptions. Evidence becomes evidence only in the context of some theoretical and conceptual framework: Only when interpreted through such a framework it is possible to claim that some data is evidence of something. And such frameworks always include background assumptions. Longino argues that these background assumptions cannot be chosen in an entirely value-free manner, as there are typically several possible and justifiable alternative frameworks available. One must choose between them, and since there is no unambiguous epistemic selection criterion, there is no way to prevent the influence of social, moral, or political values on these choices. We cannot guarantee the value-freedom of the background assumptions.

The idea of objectivity as value-freedom has also been rejected as harmful. One of the reasons for this is that in many fields of research there is a need for inherently value-laden concepts: It is not possible to study poverty, oppression, health, or well-being without using such concepts (e.g., [Dupré 2007](#)). Demanding that researchers stop using them, or claiming that a field that cannot do without such concepts cannot be objective, is untenable. The epistemic, practical, and societal value of the social and medical sciences, especially, depends on their ability to produce knowledge claims that include both empirical and moral elements ([Alexandrova 2018](#)).

Of the many objections to the demand for value-freedom, the one that has probably generated the largest amount of literature is known as the argument from inductive risk. It claims that the demand for value-freedom is irresponsible. Inductive risk is the risk of making an error in accepting or rejecting a hypothesis. It is a risk that a scientist always takes when deciding to make the inductive leap from evidence to the acceptance or rejection of a hypothesis: Is there enough evidence? Scientists also have to choose between different risks: Is it better to use a method that avoids false negative results but yields false positives, or one that avoids false positives but produces false negatives? Throughout the research process, scientists must repeatedly assess what kind of inductive risks they are willing to take and justified in taking. As many philosophers of science have pointed out, non-epistemic values must influence such decisions. This is because the predictable future use of the results matters when evaluating what risks of error are acceptable ([Douglas 2000, 2009](#); [Hempel 1965](#); [Rudner 1953](#)). To paraphrase Philip Kitcher's (2024) recent formulation of this idea, a scientist is always an ethical agent. We forbid procedures and alter study designs in science because of reasons related to research ethics, and we must similarly also make ethical judgments when evaluating what kinds of epistemic risks we can take.

It seems that if objectivity means aperspectival faithfulness to facts or value-freedom, the ideal of objectivity must be abandoned. But if this solution does not sound appealing, there are luckily many alternative accounts of objectivity to choose from.

4.4 Multiple meanings, rejections, and defences

The rejection of the two influential accounts of objectivity has led many philosophers of science to examine more closely what objectivity means in different contexts. This has led to a proliferation of recognized meanings which, first, do not reduce to either of the criticized meanings, and, second, do not seem to reduce to each other either. And as [Lloyd \(1995\)](#) has noted, we use these different meanings when ascribing objectivity to various different things, such as individuals, processes, results, or even scientific communities. When someone or something is said to be objective, this may mean, for example, that it is non-personal or impartial, or that it is produced following standardized processes, or that it has been confirmed several times using different methods, or that there is intersubjective consensus about it ([Douglas 2004](#); [Janack 2002](#); [Lloyd 1995](#); [Megill 1994](#)).

Does objectivity then mean different things in different situations? Heather [Douglas \(2004\)](#) argues that it does as follows: The concept is irreducibly complex. Nancy Cartwright, Jeremy Hardie, Eleonora Montuschi, Matthew Soleiman, and Ann C. Thresher (2022) have recently defended a similar view. They argue that objectivity is a *Ballung* concept: Abstract, imprecise, and highly dependent on context. Against such views I have argued that at least some conceptual unity can be salvaged by focusing on the function of the different meanings; at least in scientific contexts, when we say that X is objective, we argue that it is safe to rely on X, because important epistemic risks arising from our imperfections as epistemic agents have been effectively mitigated. The different meanings name different risks of this type or describe risk mitigation strategies ([Koskinen 2020, 2021](#)). Nevertheless, the proliferation of meanings has led some philosophers of science to doubt the usefulness of the

notion, echoing the arguments of earlier postcolonial and feminist thinkers who deflated the notion of any meaning other than its use in the oppression of marginalized groups.

Ian Hacking (2015) and Matthew J. Brown (2019) have explicitly suggested that we should drop the notion. Hacking (2015) claimed it to be an “elevator word.” Other such words include “real,” “true,” and “factual”—they are used to elevate the discussion to a higher level, distancing from the actual matter at hand. Instead of simply stating that the cat is on the mat, we say it is true that the cat is on the mat (Hacking 1999). According to Hacking, when we say that a statement is objective, we are primarily expressing our own attitudes toward it: We are attempting to elevate it, perhaps above criticism. Hacking sought to avoid using such words and considered the use of, and discussion about, the notion of objectivity in philosophy of science to be unnecessary. Brown (2019) has arrived at a similar conclusion because he finds the notion to be difficult or impossible to separate from the demand for value-freedom, which he rejects by invoking the argument from inductive risk. All other proposed accounts of objectivity (including the ones discussed in the next section) are in his view unsatisfactory. They do not seem to capture what we mean by “objectivity” in science, nor are they interconnected: The concept brings together a large number of disparate virtues to be cultivated and vices to be avoided in scientific theories, methods, communities, and results. If objectivity means value-freedom, it is a harmful idea, and if it means everything else that it has been claimed to mean, it is inconsistent and lacks substance. Either way, it would be best to abandon the notion altogether.

It is interesting to note that similar proposals have been widely rejected in feminist philosophy. Although MacKinnon (1987) argued that when translated into practice, objectivity simply means objectification, dehumanization, and oppression, many feminist philosophers of science have challenged this view. Mary Hawkesworth (1994) has argued that the necessary connection between objectivity and objectification on which MacKinnon’s argument rests is untenable. While we have good reasons to question the ways in which human subjects, for instance, have often been treated in science in the name of objectivity, striving for objectivity does not inevitably lead to objectification. Following similar lines, Sally Haslanger (2002) questions MacKinnon’s conclusions by distinguishing between “assumed objectivity” and “genuine objectivity.” Many feminist philosophers today agree with Sandra Harding (2015) that the notion is simply too powerful to be discarded. It is better to demonstrate how it is often misused and how feminist and postcolonial research can increase the objectivity of science. This has led to the development of accounts of objectivity that embrace the unavoidability and importance of values in science.

4.5 Value-laden objectivity

Feminist philosophers of science have developed several accounts of objectivity which often focus on different things to which we ascribe objectivity and are compatible with each other. These accounts share the premise that science is inevitably value-laden and that values can have a legitimate positive role in research (see Crasnow 2013; Scheman 2011; Wylie 2015). Before concluding this chapter, I will outline two particularly influential ones: Longino’s account of the objectivity of research communities and Harding’s distinction between weak and strong objectivity.

Longino’s (1990, 2001) account of the role of values in objective research is mostly instrumental, but nevertheless essential. As noted, she argues that the value-freedom of background assumptions in science cannot be guaranteed, and that it is therefore not possible for an individual researcher to ensure that value-laden assumptions do not bias their research in epistemically harmful ways. However, researchers do not work alone; science is a fundamentally social activity, and objectivity, according to Longino, is primarily a property of scientific communities. According to her, the best way to counter the potential distorting effects of value-laden background assumptions in science is to promote effective critical discussions and debates within scientific communities. An objective community is

diverse: Its members do not agree on value questions. This ensures, as well as possible, that even if an individual might be blind to the distorting effects of their value-laden background assumptions, someone else in the community is likely to notice them. Cultivating diverse values in scientific communities also effectively mitigates against the risk of collective biases that could go unnoticed in a homogeneous research community and possibly lead entire fields of research astray. In other words, even openly value-laden perspectives can be epistemically useful in science if they enrich critical discussions within scientific communities.

Longino (1990, 2001) has formulated four criteria that are based on these arguments and can be used to assess the objectivity of research communities. They include the following: (i) The existence of venues for effective criticism; (ii) the uptake of criticism: “beliefs and theories must change over time in response to the critical discourse taking place” (Longino 2001, 129); (iii) publicly recognized, shared standards for evaluation; and (iv) the “tempered equality of epistemic authority”: “the social position or economic power of an individual or group in a community ought not to determine who or what perspectives are taken seriously in that community” (Longino 2001, 131). While a research community should recognize expertise and can have good reasons for rejecting some forms of criticism, it is nevertheless important that it not only allows dissenting voices but cultivates them.

Standpoint epistemology allows a more directly positive role for values in science. It emphasizes the potential epistemic value of the knowledge that socially marginalized groups can possess and encourages participation in the development of such knowledge. Researchers should engage in the collective effort that allows the perspectives of members of marginalized groups to evolve into a shared, epistemically useful understanding of phenomena that might otherwise either be studied only from biased perspectives or remain entirely unexplored. Participation in the development of such standpoints contributes to the growth of knowledge and understanding, but it cannot be neutral or impartial; it is necessarily value-laden (Crasnow 2013; Harding 2004b; Rolin 2009; Wylie 2003).

Harding (1986, 2015) has defended an account of objectivity where claims made within a scientific framework, for instance, can at best be weakly objective; they have been tested in the ways known in that field, and from its perspective. But it is only when a claim has been tested and accepted in many, possibly even all situated standpoints with their own epistemic and value commitments that it can be considered strongly objective. The latter form of objectivity is obviously the more demanding and rigorous one, even though it acknowledges the epistemic value of the knowledge of socially marginalized groups and communities and gives a positive role for values in knowledge production.

4.6 Uncharted issues and missing connections

The notion of objectivity is used beyond science, and discussions about objectivity also go on in other areas of philosophy, not merely in philosophy of science. It would be useful to pay more attention to the similarities and dissimilarities between the ways in which the language of objectivity is used in different contexts. In discussions of objectivity in the philosophy of law, for instance, numerous similarities between the meanings of objectivity in science and in law have been identified, but clear differences also exist. On the one hand judges, just like scientists, are trusted to follow procedures that are assumed to promote objectivity by controlling against human biases. On the other hand, questions about whether law is or can be metaphysically objective, and questions of semantic objectivity, that is, whether statements made in some domain have truth values, are clearly more central in the philosophy of law than in the philosophy of science (Kramer 2007; Leiter 2002; Villa Rosas and Fabra-Zamora 2022).

Two missing connections that could be relevant in the discussions of values and objectivity in science relate to ethics and aesthetics. First, ethics in particular has produced a vast amount of (secular)

literature, from Moore to Rorty, Parfit, and onward, on the question whether there are objective moral values, and if so, what these are. In aesthetics there has been a similar if smaller-scale discussion about the objectivity of aesthetic values and value judgments. It is surprising how rarely this body of literature has been referenced in the discussions in philosophy of science about what kinds of values should be accepted in science. And if we take this literature into account when talking about objectivity, it at least becomes difficult to claim (like Brown does) that the notion of objectivity always invokes the idea of value-freedom.

Second, the literature on objectivity and values in science has thus far focused on moral, social, and political values, mostly leaving aside other non-epistemic values, particularly aesthetic ones. There is some interesting literature on aesthetic values in science (see [Ivanova and French 2020](#); [McAllister 2002](#)), but its connections to the literature discussed in this chapter (and book) are few and far between. This is unfortunate, since the focus on moral, social, and political values has led to arguments whose relevance is clearest in the kind of fields that inform decision-making and the kind of research where the future use of the results is somewhat predictable. It is difficult to argue that political values would generally play a significant role in mathematical research, for example. And the argument from inductive risk does not seem particularly relevant when examining basic research in theoretical physics if it is impossible to predict whether the results will ever be applied anywhere. Aesthetic values, on the other hand, can have a substantial impact on such research, and their relationship to epistemic values is unclear. On the one hand, the beauty of a theory can be misleading, but on the other, the beauty of a proof can enhance its intelligibility, and it is unclear whether the simplicity of a theory should be considered an epistemic or an aesthetic value, or both. As Longino has pointed out, a researcher's political preferences can influence which epistemic values they choose to prioritize. Similarly, aesthetic values and preferences can come into play in science and permeate the background assumptions on which research is based. How does this affect the objectivity of the research?

Acknowledgments

I am grateful to Saana Jukola, Kristina Rolin, and the editors of this volume for their helpful comments. This work was supported by the Research Council of Finland, grant No. 349051, and the SOCRATES Center for Advanced Studies at Leibniz University Hannover.

Note

- 1 I follow Julian Reiss and Jan Sprenger ([2020](#)) in using this name for this account of objectivity.

References

- Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *British Journal for the Philosophy of Science* 69: 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Biddle, Justin B. 2013. "State of the Field: Transient Underdetermination and Values in Science." *Studies in History and Philosophy of Science* 44 (1): 124–33. <http://dx.doi.org/10.1016/j.shpsa.2012.09.003>.
- Brown, Matthew J. 2019. "Is Science Really Value Free and Objective? From Objectivity to Scientific Integrity." In *What Is Scientific Knowledge?*, edited by Kevin McCain and Kostas Kampourakis. Routledge.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Cartwright, Nancy, Jeremy Hardie, Eleonora Montuschi, Matthew Soleiman, and Ann C. Thresher. 2022. *The Tangle of Science*. Oxford University Press.
- Crasnow, Sharon. 2013. "Feminist Philosophy of Science. Values and Objectivity." *Philosophy Compass* 8 (4): 413–23. <https://doi.org/10.1111/phc3.12023>.

- Daston, Lorraine. 1994. "Baconian Facts, Academic Civility, and the Prehistory of Objectivity." In *Rethinking Objectivity*, edited by Allan Megill. Duke University Press.
- Daston, Lorraine, and Peter Galison. 2007. *Objectivity*. Zone.
- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2004. "The Irreducible Complexity of Objectivity." *Synthese* 138 (3): 453–73. <https://doi.org/10.1023/B:SYNT.0000016451.18182.91>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Dupré, John. 2007. "Fact and Value." In *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, John Dupré, and Alison Wylie. Oxford University Press.
- Elliott, Kevin C. 2025. "Characterizing the Value-Free Ideal: From a Dichotomy to a Multiplicity." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw, and Carla Fehr. Boston Studies in the Philosophy and History of Science. Springer.
- Fanon, Frantz. 1961/2004. *The Wretched of the Earth*. Translated by Richard Philcox. Grove Press.
- Fine, Arthur. 1998. "The Viewpoint of No-One in Particular." *Proceedings and Addresses of the American Philosophical Association* 72 (2): 9–20.
- Hacking, Ian. 1999. *The Social Construction of What?* Harvard University Press.
- Hacking, Ian. 2015. "Let's Not Talk about Objectivity." In *Objectivity in Science. New Perspectives from Science and Technology Studies*, edited by Flavia Padovani, Alan Richardson, and Jonathan Y. Tsou. Springer.
- Haraway, Donna. 1988. "Situated Knowledges. The Science Question in Feminism and the Privilege of Partial Perspective." *Feminist Studies* 14 (3): 575–99.
- Harding, Sandra. 1986. *The Science Question in Feminism*. Cornell University Press.
- Harding, Sandra. 2004a. "A Socially Relevant Philosophy of Science? Resources from Standpoint Theory's Controversiality." *Hypatia* 19 (1): 25–47. <https://doi.org/10.1111/j.1527-2001.2004.tb01267.x>.
- Harding, Sandra. 2004b. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?" In *The Feminist Standpoint Theory Reader*, edited by Sandra Harding. Routledge.
- Harding, Sandra. 2015. *Objectivity and Diversity. Another Logic of Scientific Research*. University of Chicago Press.
- Haslanger, Sandra. 2002. "On Being Objective and Being Objectified." In *A Mind of One's Own*, edited by Louise Antony and Charlotte Witt. Routledge.
- Hawkesworth, Mary E. 1994. "From Objectivity to Objectification: Feminist Objections." In *Rethinking Objectivity*, edited by Allan Megill. Duke University Press.
- Hempel, Carl G. 1965. "Science and Human Values." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.
- Hillgardt, Hannah. 2022. "Looking Beyond Values: The Legitimacy of Social Perspectives, Opinions and Interests in Science." *European Journal for Philosophy of Science* 12 (4): 1–20. <https://doi.org/10.1007/s13194-022-00490-w>.
- Hoyningen-Huene, Paul. 2023. "Objectivity, Value-Free Science, and Inductive Risk." *European Journal for Philosophy of Science* 13 (1): 1–26. <https://doi.org/10.1007/s13194-023-00518-9>.
- Intemann, Kristen. 2010. "Twenty-Five Years of Feminist Empiricism and Standpoint Theory. Where Are We Now?" *Hypatia* 25 (4): 778–96. <https://doi.org/10.1111/j.1527-2001.2010.01138.x>.
- Ivanova, Milena, and Steven French, eds. 2020. *The Aesthetics of Science: Beauty, Imagination and Understanding*. Routledge.
- Janack, Marianne. 2002. "Dilemmas of Objectivity." *Social Epistemology* 16 (3): 267–81. <https://doi.org/10.1080/0269172022000025624>.
- John, Stephen. 2021. *Objectivity in Science*. Cambridge University Press.
- Kitcher, Philip. 2024. "The Scientist, Qua Scientist, Is an Ethical Agent." *Filozofia* 79 (3): 231–43. <https://doi.org/10.31577/filozofia.2024.79.3.1>.
- Koskinen, Inkeri. 2020. "Defending a Risk Account of Scientific Objectivity." *The British Journal for the Philosophy of Science* 71 (4): 1187–207. <https://doi.org/10.1093/bjps/axy053>.
- Koskinen, Inkeri. 2021. "Objectivity in Contexts: Withholding Epistemic Judgement as a Strategy for Mitigating Collective Bias." *Synthese* 199: 211–25. <https://doi.org/10.1007/s11229-020-02645-9>.
- Kramer, Matthew H. 2007. *Objectivity and the Rule of Law*. Cambridge University Press.
- Kuhn, Thomas. 1977. "Objectivity, Value Judgement, and Theory Choice." In *The Essential Tension. Selected Studies in Scientific Tradition and Change*. University of Chicago Press.

- Leiter, Brian. 2002. "Law and Objectivity." In *The Oxford Handbook of Jurisprudence & Philosophy of Law*, edited by Jules Coleman and Scott J. Shapiro. Oxford University Press.
- Levi, Isaac. 1960. "Must the Scientist Make Value Judgements?" *Journal of Philosophy* 57: 345–57.
- Lloyd, Elisabeth A. 1995. "Objectivity and the Double Standard for Feminist Epistemologies." *Synthese* 104 (3): 351–81. <https://doi.org/10.1007/BF01064505>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Longino, Helen E. 1996. "Cognitive and Non-Cognitive Values in Science. Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson. Kluwer Academic Publishers.
- Longino, Helen E. 2001. *The Fate of Knowledge*. Princeton University Press.
- MacKinnon, Catharine A. 1987. *Feminism Unmodified: Discourses on Life and Law*. Harvard University Press.
- McAllister, James W. 2002. "Recent Work on Aesthetics of Science." *International Studies in the Philosophy of Science* 16 (1): 7–11. <https://doi.org/10.1080/02698590120118783>.
- Megill, Allan. 1994. "Introduction. Four Senses of Objectivity." In *Rethinking Objectivity*, edited by Allan Megill. Duke University Press.
- Montuschi, Eleonora. 2003. *The Objects of Social Science*. Continuum.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford University Press.
- Reiss, Julian, and Jan Sprenger. 2020. "Scientific Objectivity." In *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/>.
- Rolin, Kristina. 2009. "Standpoint Theory as a Methodology for the Study of Power Relations." *Hypatia* 24 (4): 218–26. <https://doi.org/10.1111/j.1527-2001.2009.01070.x>.
- Rooney, Phyllis. 1992. "On Values in Science. Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1): 13–22. <https://doi.org/10.1086/psaprocbienmeetp.1992.1.192740>.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgements." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Said, Edward W. 1978. *Orientalism*. Pantheon Books.
- Scheman, Naomi. 2011. "Epistemology Resuscitated: Objectivity as Trustworthiness." In *Shifting Ground: Knowledge and Reality, Transgression and Trustworthiness*. OUP.
- Solomon, Miriam. 2012. "Socially Responsible Science and the Unity of Values." *Perspectives on Science* 20: 331–38. https://doi.org/10.1162/POSC_a_00069.
- Toole, Briana. 2022. "Objectivity in Feminist Epistemology." *Philosophy Compass* 17 (11): e12885. <https://doi.org/10.1111/phc3.12885>.
- Villa Rosas, Gonzalo, and Jorge Luis Fabra-Zamora. 2022. "Introduction: The Meanings of 'Objectivity'." In *Objectivity in Jurisprudence, Legal Interpretation and Practical Reasoning*, edited by Gonzalo Villa Rosas and Jorge Luis Fabra-Zamora. Edward Elgar Publishing.
- Ward, Zina. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Weber, Max. 1904. "Die 'Objektivität' Sozialwissenschaftlicher und Sozialpolitischer Erkenntnis." *Archiv für Sozialwissenschaft und Sozialpolitik* 19 (1): 22–87.
- Williams, Bernard. 1985. *Ethics and the Limits of Philosophy*. Harvard University Press.
- Wylie, Alison. 2003. "Why Standpoint Matters." In *Science and Other Cultures. Issues in Philosophies of Science and Technology*, edited by Robert Figueroa and Sandra Harding. Routledge.
- Wylie, Alison. 2015. "A Plurality of Pluralisms: Collaborative Practice in Archaeology." In *Objectivity in Science: New Perspectives from Science and Technology Studies*, edited by Flavia Padovani, Alan Richardson, and Jonathan Y. Tsou. Springer.

5

ARGUMENTS AGAINST THE VALUE-FREE IDEAL

Kevin C. Elliott

5.1 Introduction

The value-free ideal (VFI) has been a central organizing principle for the contemporary literature on values and science. Scholars have been discussing whether or not science should be “value free” for hundreds of years ([Proctor 1991](#)), but the recent literature on values and science has formed largely in reaction to a formulation of the VFI elucidated by Heather [Douglas \(2009\)](#) based on her historical analysis of twentieth-century debates in the philosophy of science. She describes the VFI as follows: “the value judgments internal to science, involving the evaluation and acceptance of scientific results at the heart of the research process, are to be as free as humanly possible of all social and ethical values” ([Douglas 2009](#), 45; see also [Douglas and Branch 2024](#)). The precise nature of values continues to be debated (see [Elliott and Korf 2024](#)), but for the purposes of this chapter, values can generally be regarded as desirable things or beliefs/attitudes about those desirable things. Therefore, social and ethical values would be things like public health, justice, and environmental sustainability.

It is important to emphasize several aspects of this statement of the VFI. First, it is a normative statement and not merely a descriptive one; rather than claiming that values *do not* play a role in science, it claims that they *ought not* influence science. Additionally, the VFI focuses on the *internal* aspects of science. There can be room for disagreement about precisely which aspects of science count as internal as opposed to external (see [Elliott 2025](#)), but the internal/external distinction is meant to track something like the distinction between the context of discovery and the context of justification ([Brown 2020](#)). In other words, on this account of the VFI, values can play a legitimate role in deciding what scientific hypotheses or theories to investigate, but they are not supposed to influence assessments of whether those hypotheses or theories are justified. Finally, the VFI does not exclude *all* values from influencing the internal aspects of science; rather, it excludes non-epistemic or contextual values. Proponents of the VFI are willing to allow so-called epistemic values (e.g., explanatory power, scope, fruitfulness) to influence scientific reasoning (see, e.g., [McMullin 1982](#)). The main point of the VFI is that scientific justification is supposed to be an epistemic matter; therefore, only epistemic considerations are relevant to it ([Sheykh-Rezaee and Bikaraan-Behesht 2023](#)). (Of course, opponents of the VFI tend to argue that scientific justification is *not* a purely epistemic matter; see [Lusk and Elliott 2022](#).)

This VFI has come under severe criticism from recent philosophers of science, such that the majority of scholars working on this topic no longer consider the VFI to be tenable (see, e.g., [Brown](#)

2024; Douglas and Branch 2024). This chapter provides an overview of the major arguments that have been leveled against the VFI and concludes that it is very difficult to maintain the VFI in its traditional form (see Douglas 2009). It is possible that alternative characterizations of the VFI could be developed and defended (Elliott 2025), but any plausible characterization of the VFI would have to allow for values to play significant roles in scientific reasoning and practice. Thus, moving forward it seems more fruitful to focus on exploring the best ways to manage value influences in science responsibly rather than debating the VFI.

5.2 Arguments against the VFI

This section examines six arguments that challenge the VFI: The “gap,” “inductive risk,” “conceptual,” “contingency,” “aims,” and “blurred boundaries” arguments.¹

5.2.1 The gap argument

The gap argument, which builds on the work of Helen Longino (1990), rests on the notion that there are always evidential gaps between the available data and the conclusions that scientists draw, and those gaps have to be filled by background assumptions. Longino argued that those background assumptions are often value-laden. The concept of value-ladenness can be used in different ways (Ward 2021), but Longino seemed to mean that those background assumptions were partly guided or influenced by values. For example, feminist anthropologists in the 1980s argued that even though the available evidence did not settle whether hunting activities or gathering activities played a more pivotal role in human evolution, anthropologists’ value-laden background assumptions generally inclined them toward focusing on allegedly male hunting activities (Zihlman 1985).

Taken on their own, these observations establish only that values *do* play a role in scientific reasoning, not that they can *legitimately* influence scientific reasoning (Intemann 2005). Thus, Douglas (2016) has labeled this argument as the “Descriptive Challenge” to the VFI. Proponents of the VFI could respond to this descriptive argument by insisting that scientists should try to resist the influences of values by suspending their judgment and not drawing conclusions until their background assumptions can be justified solely on the basis of evidence and epistemic values. Nevertheless, Justin Biddle (2013) has argued that it would be highly impractical for scientists to operate in this way. He notes that scientists often face situations of transient underdetermination, meaning that evidence and epistemic values are insufficient to determine which background assumptions to adopt. Moreover, he argues that it does not make sense for scientists to suspend their judgment in all these situations; it is better for them to depend on non-epistemic values in order to resolve these pervasive situations of uncertainty.

5.2.2 The argument from inductive risk (AIR)

The precise definition of the argument from inductive risk (AIR) is subject to debate,² but its basic idea is that scientists have ethical responsibilities to consider the social effects of how they respond to inductive risk (i.e., the risk of making a mistake when accepting or rejecting a hypothesis; Elliott and Richards 2017a), and this requires bringing social and ethical values into scientific reasoning. This argument has roots in the work of Churchman (1948) and Rudner (1953), was developed in a particularly influential way by Heather Douglas (2000, 2009, 2021), and has been formalized by Havstad (2022) and Brown and Stegenga (2023). In Heather Douglas’s (2000) classic account of the AIR, she illustrates the argument by considering a case where scientists are assessing the evidence that a potentially toxic chemical causes cancer. She points out that the scientists face the risk of drawing a

false positive conclusion (i.e., claiming that the chemical is carcinogenic when it is actually safe) or a false negative conclusion (i.e., claiming that the chemical is safe when it is actually carcinogenic). She argues that scientists have ethical responsibilities to demand more evidence when they can see that the social costs of drawing a false positive conclusion are likely to be high and the social costs of drawing a false negative conclusion are likely to be low, whereas scientists should demand less evidence when the social costs of drawing a false positive conclusion are lower and the social costs of drawing a false negative conclusion are higher. Importantly, [Douglas \(2000\)](#) points out that scientists make choices throughout the course of inquiry (e.g., about how to collect, analyze, and interpret data) that influence their potential to draw false positive or false negative conclusions, and so the ethical responsibility to grapple with inductive risk permeates their work.

This argument has attracted extensive scrutiny (see [Brown 2024, 2026](#); [Elliott 2022](#) for more detailed discussion of the debates about it). Perhaps the most influential response to this argument is that scientists can “defer” value-laden decisions about responding to inductive risk to others, such as policy makers (e.g., [Betz 2013](#); [Hudson 2021](#)). For example, rather than making the value-laden choice that there is sufficient evidence to conclude that a chemical is carcinogenic, scientists could simply report the evidence to policy makers and let them decide whether the evidence is sufficient. This approach could be fruitful in some cases, but it is doubtful that it would be effective for completely removing values from scientific reasoning. As Joyce Havstad and Matthew Brown (2017) have emphasized, scientific research projects incorporate a host of different choices that involve considerations of inductive risk, and it is unrealistic to think that scientists could defer them all. For example, even if scientists were able to defer the ultimate decision that there was enough evidence to declare a chemical to be carcinogenic, it is doubtful that they could defer all their other choices about how to collect, analyze, and interpret the data (which in turn make one more or less likely to draw a false positive or false negative conclusion; see [Douglas 2000](#)). Proponents of the VFI might still respond that scientists could endeavor to hedge their claims in order to make them so general and uncontroversial that there would be no risk of drawing a false conclusion (e.g., [Betz 2017](#)). So, for example, a scientist could say something like, “This particular set of data provides evidence that this chemical might possibly be carcinogenic.” However, these highly hedged claims would be of limited use for policy makers, so scientists would likely face a trade-off between the goal of providing useful information and the goal of eliminating values from their reasoning ([Brown 2024](#); [Elliott 2011](#), 55). Thus, deciding when, whether, and how much to hedge would itself be an important value-laden decision (see [Brown 2026](#)).³

5.2.3 The conceptual argument

A third argument against the VFI rests on the value-ladenness of scientific concepts. Anna [Alexandrova \(2018\)](#) has formulated this argument in a particularly compelling way by analyzing the nature of “mixed” claims, which include both descriptive and normative components. In other words, they are claims about empirical relationships, but they also include one or more variables that incorporate normative judgments. For example, consider the hypothesis that having multiple friendships is associated with greater human well-being than having no friendships. Deciding how to define well-being requires normative judgments about what well-being consists in. Therefore, according to the conceptual argument, one cannot assess the hypothesis that friendships are associated with greater human well-being without incorporating non-epistemic values in the assessment process.

It turns out that the conceptual argument applies widely across many different fields of science. Because the social sciences study human practices, they are particularly prone to employing value-laden concepts like well-being, cost of living, poverty, economic welfare, or economic inequality ([Thoma 2024](#)). All these concepts incorporate normative judgments in decisions about how to

properly define them. But even in other fields, such as biology, value-laden concepts are common (see, e.g., [Dupré 2007](#)). David [Ludwig \(2016\)](#) has extended the reach of the conceptual argument by pointing out that many concepts that do not immediately appear to be value-laden (e.g., species, memory, intelligence) end up being subject to normative decisions about how to define them in ways that best serve our practical interests (see also [Brigandt 2022](#)). More broadly, I have argued ([Elliott 2009, 2017](#)) that whenever there are social consequences associated with the use of scientific concepts, they qualify as value-laden in the sense that the choice of concepts has impacts on things we value ([Ward 2021](#)). As a result, non-epistemic values become relevant to deciding which concepts to employ and how to define them.

An influential objection to the conceptual argument is that scientists could evade it either by avoiding the use of value-laden concepts altogether or by making any claims that incorporate such concepts conditional on a particular way of defining them (see, e.g., [Alexandrova 2018](#); [Risjord et al. 2025](#); [Stegenga and Menon 2023](#)). But avoiding the use of value-laden concepts altogether seems misguided; it would make it much more difficult to study things that matter to us (e.g., well-being, poverty, or economic inequality), which generally involve such concepts (see [Dupré 2007](#)). Moreover, attempting to conditionalize the use of value-laden concepts on particular ways of defining them still forces scientists to make value-laden choices about which concepts or definitions to use. For example, an economist could carefully conditionalize their claims about a country's gross domestic product (GDP) so that they depend on a particular definition of GDP. However, the decision to use that particular definition of GDP, or the decision to use GDP rather than an alternative measure like the UN's Human Development Index (HDI), would still be a value-laden choice ([Thoma 2024](#)). One could argue that these value-laden decisions to use particular definitions or concepts would be "external" to the core of scientific reasoning, and so they would not threaten the VFI, but this response depends on drawing a sharp distinction between the internal and external aspects of science. Tim [Lewens \(2025\)](#) blurs this distinction by arguing that sometimes the case for defining concepts in particular ways is both evaluative and "internal" to science, insofar as it is grounded in scientific reasoning (see also the "blurred boundaries" argument discussed below).

5.2.4 The contingency argument

The contingency argument, which has been formalized by [Brown \(2020\)](#), rests on the notion that scientists make numerous unforced choices in the course of their work. For example, logic and evidence are often insufficient to settle how they should collect and interpret data, design studies, categorize the phenomena under investigation, develop models, and set standards of evidence for drawing conclusions. Moreover, depending on how these contingent choices are handled, they can have profound social effects ([Elliott 2017](#); [Ward 2021](#)), including "feedback effects" on which values are held and prioritized by society ([Ratti and Russo 2024](#)). Thus, given that scientists arguably have ethical responsibilities to consider the major foreseeable impacts of their actions on society ([Douglas 2009](#)), they ought to incorporate ethical and social values into their decisions about how to handle all the contingent choices associated with their work.

One way of thinking about this argument is that it is a broad umbrella that can encompass the gap argument, the AIR, the conceptual argument, and any other arguments that appeal to the ethical responsibilities of scientists to consider how the unforced choices associated with their work impact society (see also [ChoGlueck \(2018\)](#) for discussion about how these arguments relate to each other). Admittedly, [Douglas \(2000\)](#) showed that one can regard many other contingent choices in science as falling under the AIR, insofar as the other choices ultimately contribute to the downstream decision to accept or reject a hypothesis. Nevertheless, there may be some socially significant choices, such as those involved in developing and assessing models, that do not directly influence the acceptance or

rejection of hypotheses (see, e.g., [Biddle and Kukla 2017](#); [Harvard and Winsberg 2022](#)). Therefore, it may be fruitful to think of these choices as falling under the umbrella of the contingency argument but not the AIR. As with the gap, AIR, and conceptual arguments, the contingency argument can be challenged by calling for scientists to defer value-laden decisions or hedge their claims. However, a strength of the contingency argument is that it emphasizes the full range of contingent choices that scientists are forced to make (see also [Havstad and Brown 2017](#)); thus, it highlights how unrealistic it would be for scientists to fully purge their work of all value-laden decisions while still generating meaningful results.

5.2.5 The aims argument

Another argument against the VFI relies on the notion that scientific research often has non-epistemic aims in addition to epistemic ones. For example, [Elliott and McKaughan \(2014\)](#) point out that when scientists are engaged in research projects to inform regulatory decisions, they need to obtain results that are not only accurate but also usable by regulators. Therefore, when scientists assess different risk assessment methods, they need to incorporate non-epistemic values (e.g., the speed with which results can be generated and the extent to which they can be standardized) alongside epistemic ones (e.g., the accuracy of results). Similarly, Manuela Fernández Pinto and Dan Hicks (2019) argue that regulatory science has aims like protecting human health and promoting environmental justice, which means that it is appropriate to make choices about what methods to use and how to interpret ambiguous data in ways that support those values. Along similar lines, [Intemann \(2015\)](#) points out that when scientists develop and assess models, they need to consider society's non-epistemic aims as well as their epistemic aims. For example, in order to develop a climate model that predicts local precipitation patterns as accurately as possible, scientists might have to sacrifice some of the model's accuracy for predicting global temperature changes. Thus, Intemann insists that non-epistemic criteria or values are relevant for making these sorts of decisions.

The aims argument is distinct from the other arguments that can be brought under the umbrella of the contingency argument because it focuses not on scientists' general ethical responsibilities but rather on non-epistemic goals that are intrinsic to scientific practice. Thus, it is closely related to an argument that [Brown \(2020\)](#) calls the "practical reason" argument. According to Brown, scientific inquiry is a pragmatic form of inquiry that fundamentally involves taking action: "*choosing concepts, ... proposing hypotheses, ... collecting evidence, ... accepting, rejecting, inferring, asserting, or endorsing hypotheses*" ([Brown 2020](#), 65, italics added). Because scientists are involved in taking action, Brown insists that it is appropriate for them to incorporate practical reasons, such as non-epistemic values, in the course of their inquiry. Brown's practical reason argument coincides with the aims argument insofar as both appeal to the practical nature of scientific practice as a way of justifying the relevance of non-epistemic values. According to the aims argument, non-epistemic values are relevant because science sometimes incorporates non-epistemic aims, and non-epistemic values help to achieve those aims. The practical reason argument extends this point by affirming that even when science might seem to be focused on epistemic tasks, scientists are still typically engaged in practical forms of reasoning where non-epistemic values are relevant (see also [Lusk and Elliott 2022](#)).

A potential worry about the aims argument is that it seems to leave science open to epistemic corruption ([Steel 2017](#)); thus, it seems too permissive. If non-epistemic aims could potentially take priority over epistemic aims, then it seems like a scientist could justify ignoring evidence or drawing unwarranted conclusions for the sake of achieving non-epistemic aims such as making money or promoting their preferred political causes.⁴ [Lusk and Elliott \(2024\)](#) have provided several responses to this worry. One response is that epistemic aims and non-epistemic aims can often be *jointly* satisfied ([Kourany 2010](#)). Thus, one can often incorporate non-epistemic aims in scientific reasoning without

violating epistemic aims. A second response is that efforts to use science to achieve non-epistemic aims without respecting epistemic aims are typically self-defeating. In other words, the benefits of appealing to science typically rely on its authority and reliability, but that authority and reliability is likely to be lost if the epistemic aims of science are not satisfied. A final response is that even if the aims argument does not require scientists to prioritize epistemic aims over non-epistemic aims, scientists still have ethical responsibilities to be honest. Thus, if the recipients of scientific information expect epistemic aims to be prioritized in the context of scientific inquiry, then scientists have ethical responsibilities to abide by those expectations unless they can make it abundantly clear that they are prioritizing different aims. In other words, scientists should avoid “false advertising” (Carrier 2018; Wilholt 2009).

5.2.6 The blurred boundaries argument

A final argument against the VFI is what I will call the “blurred boundaries” argument. This argument rests on the fact that the VFI is typically formulated in such a way that it relies on sharp distinctions between different kinds of values and different aspects of science. For example, it allows only a narrow range of values (e.g., epistemic ones) to influence the “internal” aspects of scientific reasoning, whereas a broader range of values (e.g., non-epistemic ones) are allowed to influence the “external” aspects of science. The internal aspects of scientific reasoning are typically interpreted to involve the assessment and acceptance of hypotheses, whereas the external aspects involve activities like choosing research questions, setting ethical guidelines about how studies should be conducted, and applying scientific results to public policy.

The blurred boundaries argument questions one or more of these distinctions on which the VFI relies. As a result, the VFI becomes incoherent because the value influences that are supposed to be prohibited end up being indistinguishable from the value influences that are supposed to be allowed. Historically, the most common way to develop the blurred boundaries argument has been to challenge the distinction between epistemic and non-epistemic values. For example, Longino (1996) famously argued that there was not a convincing basis for distinguishing so-called epistemic or “constitutive” values (e.g., consistency, simplicity, and fruitfulness) from alternative feminist values (e.g., ontological heterogeneity, novelty, and applicability). Similarly, Phyllis Rooney (1992) argued that the application of supposed epistemic values is often influenced by non-epistemic values, and more recently she has pointed out that values often operate in a fuzzy “borderlands” between the epistemic and the non-epistemic (Rooney 2017).

Daniel Steel (2010) has attempted to shore up the distinction between epistemic and non-epistemic values by defining epistemic values as those that promote the attainment of truth and by clarifying that any particular value (e.g., explanatory power, simplicity, ontological heterogeneity, or novelty) can be truth-promoting in some contexts but not in others. However, the problem with Steel’s approach is that it facilitates a clear distinction between epistemic and non-epistemic values *in principle*, but it does not provide clear guidance for scientists *in practice*. As they go about their work, scientists cannot be entirely sure which values are actually truth-promoting in the context in which they are working. Some values that typically count as epistemic might not actually be truth-promoting in a particular context, and some values that typically count as non-epistemic might actually be truth-promoting. Thus, Steel’s argumentative strategy could perhaps save the VFI from this form of the blurred boundaries argument if the VFI were interpreted as describing an ideal state that ought to be achieved, but Steel’s move does not work if the VFI is to serve as a practical guideline for working scientists (Elliott 2025).

More recently, the blurred boundaries argument has been extended so that it challenges not only the epistemic/non-epistemic value distinction but also the distinction between the internal and

external aspects of science (see, e.g., [Resnik and Elliott 2023](#)). Kathleen Okruhlik (1994) pioneered this argumentative move by showing how scientists' assessments of which theory is most compelling in a particular domain depend on the range of alternative theories under consideration, which is in turn influenced by the values of those developing and proposing the theories. [Elliott and McKaughan \(2009\)](#) similarly emphasized that scientists' assessment of the evidential support for theories (which is generally regarded as an "internal" aspect of scientific reasoning) is influenced by the kinds of data that scientists collect and the kinds of theories that they propose (which are generally regarded as "external" aspects of scientific reasoning). Thus, Elliott and McKaughan argued that when values influence the external aspects of science, those influences bleed into the internal aspects of science. Along the same lines, [Holman and Bruner \(2017\)](#) have discussed the process of "industrial selection," whereby private companies selectively fund investigators who employ methods that tend to generate results that favor their preferred theories. By doing so, they can influence the overall body of evidence in a particular research area even without biasing the reasoning of any individual scientist (see also [O'Connor and Freeborn 2025](#)). Based on these observations, it seems difficult to maintain a form of the VFI that allows values to influence external aspects of science (like choosing which research questions to prioritize or which studies to pursue) while insisting that values should not influence internal aspects of science (like assessing the evidential support in favor of a theory or hypothesis).⁵

5.3 Moving forward

The upshot of these arguments is that the "traditional" VFI described by [Douglas \(2009\)](#) is very difficult to defend. As [Brown \(2024\)](#) points out, to maintain the VFI would require challenging the AIR, which appears to be valid and has been defended as sound, as well as overcoming the variety of other arguments that form "a cable of arguments pulling against the VFI" (2024, 26). Nevertheless, it is important to recognize that the VFI can be characterized in different ways, and there is currently renewed interest in developing new formulations of it that allow values to influence science but that strive to minimize those influences (see, e.g., [Menon and Stegenga 2023, 2026](#); [Sikorski 2024](#); [Stamenkovic 2024](#)). Brown rejects this move by arguing that to accept or reject the VFI is an "all-or-nothing question" (2024, 26); he insists that those who try to defend "weaker" versions of the VFI are really just discussing strategies for managing values ([Brown 2026](#)). However, this response rests on the assumption that Brown's notion of the VFI—what I have been calling the "traditional" notion developed by [Douglas \(2009\)](#)—is the only legitimate formulation of it. I think this assumption is questionable.

One can legitimately disagree about what kinds of value influences in science are the most significant or important for scientists to consider, and thus what kinds of value influences should be enshrined in the VFI ([Elliott 2025](#)). The formulation of the VFI that [Douglas \(2009\)](#) developed based on twentieth-century philosophical debates is in fact quite nuanced because it is informed by the failure of more stringent efforts to exclude values from science. As a result, one might think that Douglas's formulation of the VFI has already caved in and abandoned "real" value-freedom for science. In other words, one might insist that a truly comprehensive formulation of the VFI should prevent ethical or social values from influencing *any* part of (basic) science, or one might insist that *all* values (not just ethical and social values) should be excluded from scientific reasoning. Thus, one could argue that those who are trying to develop new formulations of the VFI are just continuing a previous tradition of gradually loosening up the standards for the VFI while insisting that some value influences should still be excluded or minimized.

Brown might reply that Douglas's formulation of the VFI obviously captures what really matters because it excludes social and ethical values from the "heart" of the research process. However, this begs the question of what should count as the heart of the research process. For example, one could

insist that values do not really enter the heart of the research process unless they play the role of scientific evidence, and [Douglas \(2009\)](#) would actually maintain this formulation of the VFI rather than rejecting it. Thus, deciding how to formulate the VFI is not a straightforward matter; it depends on assumptions about what sorts of value influences are significant enough and controversial enough to try to prohibit ([Elliott 2025](#)), and these assumptions change over time depending on contingent historical debates and concerns ([Proctor 1991](#)). For example, Philip Kitcher recently proposed this formulation of the VFI, which is quite different from the traditional formulation that most philosophers have been discussing: “In a scientist’s scheme of values, epistemic values should outrank all other values” ([Kitcher 2024](#), 233).

Nevertheless, even though decisions about how to define the VFI are contingent and subject to change, the arguments discussed in this chapter indicate that any reasonable contemporary formulation of the VFI will have to allow values to play pervasive roles in science. Therefore, even though one could, in theory, reformulate the VFI so as to make it worthy of contemporary debate, this seems likely to be more confusing than helpful, especially for those outside the philosophical community. In sum, [Brown \(2024\)](#) might be wrong *in principle* when he says that debates about the VFI are really just debates about value management, but he is probably correct *in practice* that it would be better to shift the terms of the debate in this way. Rather than framing future discussions about values and science as debates about the VFI, it seems more fruitful to focus on exploring the best ways to manage value influences responsibly (see, e.g., [Holman and Wilholt 2022](#); [Lusk 2025](#)). The approaches being promoted to try to save some form of the VFI (see, e.g., [Levy 2025](#); [Menon and Stegenga 2023, 2026](#); [Parker 2024, 2026](#); [Sikorski 2024](#); [Stamenkovic 2024](#)) could then be interpreted as strategies for responsibly managing values in science.

Acknowledgments

I am grateful to Matt Brown for very helpful comments on an earlier version of this chapter. I am also thankful for funding support I received through a Research Award from the Alexander von Humboldt Foundation and through the SOCRATES Center at Leibniz University, Hannover, which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project 470816212/KFG43.

Notes

- 1 Another potential argument against the VFI is that it rests on a problematic, noncognitivist conception of values, and so it illegitimately presumes that values cannot stand in evidential relationships with scientific claims (see, e.g., [Anderson 2004](#); [Brown 2020](#); [Clough 2008](#)).
- 2 The debate over the AIR includes what elements of science are subject to the argument as well as what range of consequences to take into account (see [Elliott and Richards 2017b](#), 263–65). I thank Arnon Levy for pointing out that my tendency in the past to refer to the AIR as the “error argument” (see [Elliott 2011, 2022](#)) could misleadingly suggest that only the consequences of error need to be taken into account, but I am not committed to that position.
- 3 Another potential response to the AIR is to argue that it is problematic for individual scientists to make idiosyncratic decisions about how to respond to inductive risk, as opposed to having groups of scientists make standardized decisions about handling inductive risk (see e.g., [John 2015](#); [Levy 2025](#)). Although this response would not technically save the VFI (because scientists would still be incorporating values in their reasoning at a group level), it could be interpreted as supporting a form of value freedom at the individual level. It seems unlikely, however, that responsible scientists could avoid all considerations of inductive risk at the individual level, in part because they would have to decide when (if ever) to depart from the group-level standards.

- 4 One response to this objection would be to insist, along the lines of Brown's practical reason argument, that there are no purely epistemic aims of science. For the sake of argument, however, I will assume that it is coherent to distinguish epistemic aims of science from non-epistemic aims.
- 5 Yet another potential way to extend the blurred boundaries argument would be to argue that it is impractical for scientists to distinguish between their different cognitive attitudes (e.g., belief, acceptance, endorsement, or pursuit) and to point out that non-epistemic values are relevant to at least some of these attitudes (see Elliott 2022, 36).

References

- Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *The British Journal for the Philosophy of Science* 69(2): 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Anderson, Elizabeth. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19 (1): 1–24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>.
- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Betz, Gregor. 2017. "Why the Argument from Inductive Risk Doesn't Justify Incorporating Non-Epistemic Values in Scientific Reasoning." In *Current Controversies in Values in Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge. <https://doi.org/10.4324/9781315639420-6>.
- Biddle, Justin B. 2013. "State of the Field: Transient Underdetermination and Values in Science." *Studies in History and Philosophy of Science* 44 (1): 124–33. <http://dx.doi.org/10.1016/j.shpsa.2012.09.003>.
- Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Brigandt, Ingo. 2022. "How to Philosophically Tackle Kinds without Talking about 'Natural kinds'." *Canadian Journal of Philosophy* 52 (3): 356–79. <https://doi.org/10.1017/can.2020.29>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Brown, Matthew J. 2026. "Recent Arguments for the Ideal of Value-Free Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Brown, Matthew J., and Jacob Stegenga. 2023. "The Validity of the Argument from Inductive Risk." *Canadian Journal of Philosophy* 53 (2): 187–90. <https://doi.org/10.1017/can.2023.37>.
- Carrier, Martin. 2018. "Identifying Agnotological Ploys: How to Stay Clear of Unjustified Dissent." In *Philosophy of Science: Between the Natural Sciences, the Social Sciences, and the Humanities*, edited by Alexander Christin, David Hommen, Nina Retzlaff, and Gerhard Schurz. Springer.
- ChoGlueck, Christopher. 2018. "The Error Is in the Gap: Synthesizing Accounts for Societal Values in Science." *Philosophy of Science* 85 (4): 704–25. <https://doi.org/10.1086/699191>.
- Churchman, C. West. 1948. *Theory of Experimental Inference*. Macmillan.
- Clough, Sharyn. 2008. "Solomon's Empirical/Non-Empirical Distinction and the Proper Place of Values in Science." *Perspectives on Science* 16 (3): 265–79. <https://doi.org/10.1162/posc.2008.16.3.265>.
- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2016. "Values in Science." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.
- Douglas, Heather. 2021. *The Rightful Place of Science: Science, Values, and Democracy: The 2016 Descartes Lectures*. Edited by Ted Richards. Consortium for Science, Policy & Outcomes.
- Douglas, Heather, and T. Y. Branch. 2024. "The Social Contract for Science and The Value-Free Ideal." *Synthese* 203: 40. <https://doi.org/10.1007/s11229-023-04477-9>.
- Dupré, John. 2007. "Fact and Value." In *Value-Free Science? Ideals and Illusions*, edited by Harold Kincaid, Alison Wylie, and John Dupré. Oxford University Press.
- Elliott, Kevin C. 2009. "The Ethical Significance of Language in the Environmental Sciences: Case Studies from Pollution Research." *Ethics, Place & Environment* 12: 157–73.

- Elliott, Kevin C. 2011. *Is a Little Pollution Good for You? Incorporating Societal Values in Environmental Research*. Oxford University Press.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin C. 2025. "Characterizing the Value-Free Ideal: From a Dichotomy to a Multiplicity." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw, and Carla Fehr. Springer.
- Elliott, Kevin C., and Rebecca Korf. 2024. "Values in Science: What Are Values, Anyway?" *European Journal for Philosophy of Science* 14: 1–24. <https://doi.org/10.1007/s13194-024-00615-3>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. "How Values in Scientific Discovery and Pursuit Alter Theory Appraisal." *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81: 1–21. <https://doi.org/10.1086/674345>.
- Elliott, Kevin C., and Ted Richards. 2017a. "Exploring Inductive Risk: An Introduction." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Elliott, Kevin C., and Ted Richards. 2017b. "Exploring Inductive Risk: Future Questions." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Fernández Pinto, Manuela, and Dan Hicks. 2019. "Legitimizing Values in Regulatory Science." *Environmental Health Perspectives* 127 (3): 035001. <https://pubmed.ncbi.nlm.nih.gov/30870036/>.
- Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32 (1): 1–31. <https://doi.org/10.1353/ken.2022.0001>.
- Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Havstad, Joyce C., and Matthew J. Brown. 2017. "Inductive Risk, Deferred Decisions, and Climate Science Advising." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Holman, Bennett, and Justin Bruner. 2017. "Experimentation by Industrial Selection." *Philosophy of Science* 84 (5): 1008–19. <https://doi.org/10.1086/694037>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Hudson, Robert. 2021. "Should We Strive to Make Science Bias-Free? A Philosophical Assessment of the Reproducibility Crisis." *Journal for General Philosophy of Science* 52 (3): 389–405. <https://doi.org/10.1007/s10838-020-09548-w>.
- Intemann, Kristen. 2005. "Feminism, Underdetermination, and Values in Science." *Philosophy of Science* 72 (5): 1001–12. <https://doi.org/10.1086/508956>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5: 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- John, Stephen. 2015. "Inductive Risk and the Contexts of Communication." *Synthese* 192 (1): 79–96. <https://doi.org/10.1007/s11229-014-0554-7>.
- Kitcher, Philip. 2024. "The Scientist, Qua Scientist, Is an Ethical Agent." *Filozofia* 79 (3): 231–43. <https://doi.org/10.31577/filozofia.2024.79.3.1>.
- Kourany, Janet A. 2010. *Philosophy of Science after Feminism*. Oxford University Press.
- Levy, Arnon. 2025. "A (Consequence-Oriented) Critique of the Argument from Inductive Risk." *Philosophical Studies* 182: 2079–99. <https://doi.org/10.1007/s11098-025-02319-7>.
- Lewens, Tim. 2025. "Towards a Direct Role for Values in the Heart of Science." *Philosophy of Science*: 1–25. <https://doi.org/10.1017/psa.2025.18>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Longino, Helen E. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson. Vol. 256. Synthese Library. Kluwer.
- Ludwig, David. 2016. "Ontological Choices and the Value-Free Ideal." *Erkenntnis* 81 (6): 1253–72. <https://doi.org/10.1007/s10670-015-9793-3>.
- Lusk, Greg. 2025. "Science, Values, and Action Guidance: Can We Stop Talking about the Value Free Ideal?" In *Values, Pluralism, and Pragmatism*, edited by Jonathan Y. Tsou, Jamie Shaw, and Carla Fehr. Springer.

- Lusk, Greg, and Kevin Elliott. 2024. "Against Global Aims for Science: Values, Epistemic Priority, and a Local Aims Approach." *Synthese* 204 (2): 47. <https://doi.org/10.1007/s11229-024-04633-9>.
- Lusk, Greg, and Kevin C. Elliott. 2022. "Non-Epistemic Values and Scientific Assessment: An Adequacy-for-Purpose View." *European Journal for Philosophy of Science* 12 (2): 35. <https://doi.org/10.1007/s13194-022-00458-w>.
- McMullin, Ernan. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (2): 2–28. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409>.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom Is Worth Pursuing." *European Journal for Philosophy of Science* 13 (4): 48. <https://doi.org/10.1007/s13194-023-00552-7>.
- Menon, Tarun, and Jacob Stegenga. 2026. "Democracy, Consensus, and the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- O'Connor, Cailin, and David Peter Wallis Freeborn. 2025. "Industrial Distraction." *Philosophy of Science*: 1–22. <https://doi.org/10.1017/psa.2025.1>.
- Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24 (sup1): 21–42. <https://doi.org/10.1080/00455091.1994.10717393>.
- Parker, Wendy S. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36. <https://doi.org/10.1017/psa.2023.107>.
- Parker, Wendy S. 2026. "An Intermediate Approach to Value Management." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Proctor, Robert. 1991. *Value-Free Science?* Harvard University Press.
- Ratti, Emanuele, and Federica Russo. 2024. "Science and Values: A Two-Way Direction." *European Journal for Philosophy of Science* 14: 6. <https://doi.org/10.1007/s13194-024-00567-8>.
- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54: 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Risjord, Mark, Kareem Khalifa, and Jared Millson. 2025. "Thick Concepts and Impartiality." *Philosophy of Science*. <https://doi.org/10.1017/psa.2025.10114>
- Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (1): 13–22. <https://doi.org/10.1086/psaprocbienmeetp.1992.1.192740>.
- Rooney, Phyllis. 2017. "The Borderlands between Epistemic and Non-Epistemic Values." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Sheykh-Rezaee, Hossein, and Hamed Bikaraan-Behesht. 2023. "Value-Free Ideal Is an Epistemic Ideal: An Objection to the Argument from Inductive Risk." *Principia: An International Journal of Epistemology* 27 (1): 137–63. <https://doi.org/10.5007/1808-1711.2023.e80487>.
- Sikorski, Michał. 2024. "Values, Bias, and Replicability." *Synthese* 203: 164. <https://doi.org/10.1007/s11229-024-04573-4>.
- Stamenkovic, Philippe. 2024. "Straightening the 'Value-Laden Turn': Minimising the Influence of Extra-Scientific Values in Science." *Synthese* 203: 20. <https://doi.org/10.1007/s11229-023-04446-2>.
- Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. <https://doi.org/10.1086/650206>.
- Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Stegenga, Jacob, and Tarun Menon. 2023. "The Difference-to-Inference Model for Values in Science." *Res Philosophica* 100 (4): 423–47. <https://doi.org/10.5840/resphilosophica2023928102>.
- Thoma, Johanna. 2024. "Social Science, Policy and Democracy." *Philosophy and Public Affairs* 52: 5–41. <https://doi.org/10.1111/papa.12250>.
- Ward, Zina B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science Part A* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science* 40: 92–101. <https://doi.org/10.1016/j.shpsa.2008.12.005>.
- Zihlman, Adrienne L. 1985. "Gathering Stories for Hunting Human Nature." *Feminist Studies* 11 (2): 365–77. <https://doi.org/10.2307/3177929>.

6

RECENT ARGUMENTS FOR THE IDEAL OF VALUE-FREE SCIENCE

Matthew J. Brown

6.1 Introduction

Historically, the common view is that science is, or ought to be, neutral or impartial with respect to social, moral, and political values, that at least as an ideal, science ought to be value-free. Since the turn of the century, philosophers of science increasingly hold that the ideal of value-free science is untenable and that science is and ought to be value-laden in some respect; some even describe this as a “general consensus” (Holman and Wilholt 2022; Lusk 2021). Elliott (2026) reviews the arguments in favor of this view. However, this consensus is not without challengers; indeed, defenses of the value-free ideal (VFI) appear with regularity. In this chapter, I will review the main lines of defense of the VFI in the recent literature and assess their adequacy, arguing that they fail to provide a convincing response to the arguments against the VFI.

6.2 The value-free ideal

The ideal of value-free science, or the **value-free ideal (VFI)**, asserts that there is a core part of science (sometimes called the “internal” or “inferential” part) where certain values (“non-epistemic values”) ought not be permitted to have an influence (Douglas 2009, Ch 3; Douglas 2016). According to the VFI, only epistemic values are allowed in the core; thus, we could call it “the ideal of epistemic purity” instead (Biddle 2013). The VFI does not necessarily apply to all parts of science; many, though not all, of its defenders acknowledge that values have a legitimate role to play in determining which questions or problems scientists pursue or how they frame hypotheses for consideration. Many also acknowledge that scientific methods must be constrained by ethical considerations such as the wellbeing of research subjects. However, when it comes to the constitution of evidence or evaluating the way that evidence tells for or against a hypothesis, only epistemic values may play a role. Scientists ought not consider non-epistemic values when making such judgments in science. Though there is room to debate which values must be proscribed from which phases of scientific inquiry to achieve the “epistemic purity” that the VFI demands, to defend the VFI is to insist that science ought to be *absolutely* value-free within those bounds.

The VFI is a normative ideal, not a descriptive claim. It is meant to govern the kinds of considerations scientists should weigh in making and justifying their decisions, actions, and claims, as well as the social conventions and institutional structures that mediate the assessment of scientific claims.

Defenders of the VFI acknowledge that values sometimes *cause* scientists to act or reason in certain ways or that values often *motivate* the way that they pursue their work. Defenders of the VFI readily admit that scientists are human and not perfect epistemic machines. What matters is that scientists *strive* for value-free reasoning and adopt practices that minimize the role of non-epistemic values as (direct or indirect) *justifying* reasons. In terms of Ward (2021), it is the role of values as “justifying reasons” that is centrally at issue in arguments for and against the VFI. As an *ideal*, the VFI is insulated from ought-implies-can arguments. It does not matter whether it is achievable; it may still be worth pursuing (see Menon and Stegenga 2023). The question that concerns us here is whether scientists *in practice* ought to be *guided* by the VFI, whether they ought to take it as a *regulative ideal* of their activity, or whether they ought to *strive* toward value-freedom.

An adequate critique of the VFI must be a normative argument that, all things considered, it is not desirable for scientists to pursue nor achieve the VFI. Such critiques must show that it is normatively legitimate, even required, for scientists to consider values in the ordinary course of their work, in some core or internal part of science. Such critiques are reviewed in Elliott’s chapter in this volume (Elliott 2026). An adequate defense of the VFI likewise must demonstrate that the VFI is both normatively desirable and provides some actionable guidance for scientific practice.

6.3 Distinguishing the VFI and the value-management questions

The literature on science and values at present is large and complex, but two questions are central: (i) **The VFI question**—normatively, should science be value-free, i.e., ought scientists to regulate their practice by aiming at value-freedom? (ii) **The value-management question**—assuming that science is *not* value-free, how ought the role of values in science be managed, that is, when and how should values be permitted to operate in science? (See Silk (2018) on this distinction. The “value-management question” has also recently been called “the new demarcation problem” by Holman and Wilholt (2022).)

The two questions are sometimes conflated, giving rise to various confusions. One confusion is that there is a *spectrum* of positions on values in science, with the VFI at one extreme, and some sort of radical value-ladenness at the other. This characterization is a mistake, however, given that the VFI is an all-or-nothing affair—any role, to any degree, for social and ethical values in the internal phases of scientific reasoning means that science is not *free* from values. And this matches what most defenders of the VFI intend; they are defending an all-or-nothing purity claim: no non-epistemic values allowed in the internal phases of science. While different interpretations of key elements of the VFI (e.g., “epistemic values” and “internal phases”) can lead to different interpretations or versions of the VFI, that is very different from a continuous spectrum between value-freedom and its denial. A related confusion is to treat more conservative positions on the second question as if they were partial defenses of the VFI. Often, the best work that claims to be defending the VFI is better understood as answering the value-management question. One might find the latter question more interesting, but it only makes sense to address it once the VFI has been rejected.

6.4 Arguments for the value-free ideal

The most influential and strongest argument against the VFI in the recent literature is the *argument from inductive risk* (AIR). The contemporary form of the AIR is due to the groundbreaking work of Heather Douglas (2000, 2009). Brown and Stegenga (2023), building on Havstad (2022), have provided a reconstruction of Douglas’s (2000) AIR that shows that the argument is valid and, arguably, sound. Many defenses of the VFI respond to the AIR, and so their adequacy in this respect will be one theme in this chapter. But there are many other relevant arguments against the VFI (see Elliott 2026), many of which have received little attention by defenders of the VFI.

6.4.1 Deferring and hedging

One classic defense of the VFI is the *deferred decision response* or *hedging response*. Classic examples of this response can be seen in Jeffrey (1956), Mitchell (2004), Betz (2013), and Edenhofer and Kowarsch (2015) (although, see Harvard and Winsberg (2022) for an alternative interpretation of Jeffrey that denies that he accepts the VFI). These responses argue that, though value judgments are necessary in order to reach conclusions about hypothesis acceptance or rejection, it is not part of the core of science to reach such conclusions. Instead, scientists can defer these decisions to the context of application, providing all the relevant information (such as probabilities of the hypothesis given the evidence) needed. The conclusions that scientists should reach are suitably hedged so as to avoid inductive risks entirely.

Several lines of response have been pursued to this type of argument. The first points toward a regress of inductive risks, showing that scientists cannot in fact defer or hedge effectively and responsibly. This regress goes in two directions, *downstream* or *upstream* of hypothesis evaluation.

The *downstream regress* of inductive risks is pointed out already by Rudner (1953). Rudner anticipated the deferred decision response and replied that inductive risks are also present in whatever information is presented to decision-makers. If instead of asserting H , the scientists assert $P(H) = p$, there are inductive risks here as well. Steel (2016) calls this “second-order uncertainty.” The downstream regress has quantitative and qualitative aspects. Quantitatively, the assertion $P(H) = p$ is not itself completely certain, but itself has inductive risks associated. Likewise, $P(P(H) = p) = p'$ and so on. Though the practical significance of these risks may decrease as one travels down the regress, certainty is never reached, and so inductive risk is never eliminated. At a practical level, whether there are foreseeable non-epistemic consequences associated with decisions about N-order uncertainties cannot be determined in advance, *a priori*; it is itself a context-specific value judgment. Qualitatively, there are multiple options for models and methods of estimating probabilities that lead to different probability ascriptions and so different conclusions. In a related but more technical way, Steel (2015) shows that a Bayesian analysis of confirmation remains subject to the AIR, since the determination of prior probabilities and likelihoods involves a decision to accept a probability model, a decision internal to science that has its own inductive risks.

There is also an *upstream regress* of inductive risks. The phenomenon of inductive risk applies not only to the final judgment about hypothesis acceptance. There are many intermediate judgments in scientific inquiry that are themselves uncertain, with potentially significant non-epistemic consequences. Havstad and Brown (2017) follow Douglas (2009) and Winsberg (2012) in arguing that these upstream decisions are too many, and too complex, to be dealt with by deferred decisions. What’s more, the consequences of those upstream decisions may not align with those of the downstream decisions, but may bring in unique considerations. One might defer a small subset of value-laden decisions, but not all.

Frisch (2020) shows that the decisions cannot be legitimately deferred or hedged based on a principle articulated by Elliott (2011): the *no-passing-the-buck principle*. Elliott points out that it can be harmful for scientists to withhold judgment on matters when their technical judgment can help inform decision-makers. As Frisch argues, “One way in which scientists may violate the no-passing-the-buck principle is to commit only to a strongly hedged variant of a hypothesis P while withholding judgment on P itself” (Frisch 2020, 983). There is a kind of trade-off between informativeness and certainty, which is itself a kind of epistemic risk different from but relevant to inductive risks.¹ Hedging gets you certainty at the cost of informativeness, but science must be *informative* and *policy-relevant* in order to be consequential to decision-makers and thus to fulfill the social role that gives it social and political weight (Elliott 2011; Frisch 2020; Steel 2016; Steele 2012; see Menon and Stegenga 2023). It is not an either-or proposition (hedged or informative), but rather a trade-off that

must be judged by using values. Hedging allows one to mitigate inductive risks at the cost of limiting the social value of science. No one ought to deny that there are contexts where hedging or deferring to *some* extent turns out to be best; what opponents of the VFI argue is that non-epistemic consequences must be considered in determining whether and how much to hedge, and how much inductive risk must be embraced. They rightly deny that it is *always* best, in every context, for scientists to present the *most strongly* hedged claims. Finding the right balance depends in part on non-epistemic values, thus undermining the argument against values in science.

Several recent defenses of the VFI attempt to revive the deferred decision or hedging response (Cassini 2022; Henschen 2021; MacGillivray 2019). Most of these attempts do not appropriately engage with the past criticisms of similar versions of this response.

For instance, Henschen (2021) argues that the regress arguments against the deferred decision response fail, because in scientific practice, scientists typically don't consider higher-order uncertainties (8–11). Henschen is aware of both the upstream and downstream regresses as presented by Rudner and Douglas, respectively. His argument tries to reduce the upstream regress of inductive risks—which concerns decisions like the weighing of inductive risks in determining evidence prior to considering how the evidence bears on the hypothesis—to the downstream regress of inductive risks—which concerns second-order uncertainty, the inductive risks of accepting $P(H) = p$. Henschen writes, “the problem with Douglas’s suggestion is that the inductive risk that is present at the three lower stages adds up to the total risk of the null hypothesis in accordance with the laws of probability” (2021, 7).

This argument fails because Henschen confuses inductive risks with uncertainty (2021, 7). Imagine an experiment meant to assess whether chemical X is carcinogenic; there may be some uncertainty introduced in the measurement of cancers, uncertainty introduced in the modeling of dose-response relationships, and uncertainty associated with the calculation of the mean effect. If these uncertainties are quantifiable, straightforward mathematics (error analysis) allows us to aggregate them into an overall probability of error. Risk (in the sense that is relevant to decisions under inductive risk), on the other hand, involves both uncertainty or probability of error *and* the (dis)value of the outcome. If the benefit or harm of an outcome is quantifiable, then risks can be aggregated using an *expected utility calculation*, but only if one adds to the mere uncertainties the possible outcomes, the probabilities of the outcomes conditional on the options being decided, and the values of the outcomes—in other words, only if one reasons in a value-laden way (Winsberg 2024). And if the relevant outcomes associated with measurement uncertainty are different from modeling uncertainty, and so on for the many different judgments made in the course of scientific inquiry, the complexity of the relevant calculation grows exponentially, such that it cannot be deferred or hedged. What's more, the relevant values may not be quantifiable or commensurable. Henschen's response to the downstream regress issue is also inadequate. Recall, there are both quantitative and qualitative aspects to the problem; he ignores the qualitative aspect of the problem entirely.

This section has reviewed classic and recent versions of the deferred decision response. The response has serious flaws, and recent versions have done little to overcome them.

6.4.2 The democratic defense

Another major argument in defense of the VFI is that value-laden science is incompatible with the requirements of liberal democracy. This has been called “the *democratic defense*” and “the *political legitimacy argument*.” The argument is that we must uphold the VFI in order for science to play a legitimate role in democratic governance, which we cannot really do without (see Douglas 2009, 8 and chap. 2). For instance, in one of the most interesting versions of this argument, Bright (2018) reconstructs W.E.B. Du Bois as arguing that, if scientists are not seen to be following the VFI, they

will lose public support and trust, and this will undermine the attempt to use science to forward socially valuable goals. (This oft-repeated empirical claim about the public is doubtful in the face of the evidence presented by [Hicks and Lobato \(2022\)](#).) Likewise, referring to the AIR and the older underdetermination arguments, [Lusk \(2021, 103\)](#) claims that arguments against the VFI fail to address concerns about political legitimacy (see also [Betz 2013](#); [Carrier 2022](#); [Holman and Wilholt 2022](#); [Kappel 2014](#); [Kappel and Zahle 2019](#)).

This is one of the most serious concerns about value-laden science. It is useful to look at [Lusk's \(2021\)](#) version of the argument (I have omitted one unnecessary and one redundant premise from Lusk's original presentation):

2. [Legitimacy Premise] No set of non-epistemic values should have an undue influence on coercive democratic political decisions.
3. [Infiltration Premise] If non-epistemic values play a role in the empirical justification of political decisions, then those values have an undue influence [...]
- C. Therefore, it is not the case that non-epistemic values should play a role in empirical justification in democratic decision-making. (104)

If this argument is correct, one important background justification of the AIR, that science has a clear and significant social impact because of its role in policymaking, becomes a reason instead in support of the VFI. However, each premise of this argument is shaky. [Lusk \(2021\)](#) concentrates on the third premise (infiltration), arguing that it is possible for non-epistemic values to have a role in justification without “undue influence.” Also, against the legitimacy premise, it seems clear that *some* non-epistemic values have a legitimate role to play in democratic decisions, e.g., the ethical values constitutive of science, the political values constitutive of liberal democracy, or, in some cases, the values of the majority.

The political legitimacy argument and the arguments against the VFI cannot both be sound, but the former is on much shakier ground than the latter. As the political legitimacy argument is typically not accompanied by any direct refutation of arguments against the VFI, one suspects something has gone wrong. For instance, [Wilholt \(2023\)](#) responds to Du Bois's claim (via [Bright 2018](#)) that science must aim solely to discover the truth by pointing out that the AIR does not deny that science aims to discover the truth; rather, the AIR shows that, in pursuing the truth, one meets IR that must be managed, that cannot be settled by the aim of truth. The VFI cannot guide scientific investigation in a trustworthy fashion by recommending negligence concerning these risks.

In my view (following Lusk), although many of these arguments are presented as defenses of the VFI, they do not really amount to full defenses despite the intentions of the authors. Rather, they present a challenge that accounts of value-laden science need to address. In other words, they conflate the VFI question with the value-management question; their real value is in challenging us to think carefully about the latter in the context of the political roles of science. And the literature has risen to the challenge, shifting a great deal of attention in recent discussions to matters of democratic values and the political legitimacy of value-laden science (e.g., [Alexandrova and Fabian 2022](#); [Boulcault and Schroeder 2021](#); [Douglas 2005, 2012](#); [Kitcher 2011](#); [Lusk 2021](#); [Fernández Pinto and Hicks 2019](#); [Schroeder 2021](#)).

6.4.3 Ideal theory

Another approach to defending the VFI is to retreat from the messy world of scientific practice and the science-policy or science-society interface to the realm of epistemic ideals. This sort of retreat raises worries about whether the resulting argument will have any purchase within scientific practice,

but we should not dismiss these arguments out of hand. We should ask, instead, can the ideal theory thus devised shed any light or provide any guidance to scientific practice, or our evaluation of it?

Hudson (2016) uses an ideal theory approach in another attempt to revive the deferred decision response. He acknowledges the problems that have been raised with this response, but he dismisses these concerns with the claim that VFI is an “epistemic ideal” that sets out what is “epistemically preferable.” This does not really address the terms of the argument about what is *legitimate* in actual scientific reasoning. That is, rather than address objections head-on and tell us what scientists could do to defer effectively and responsibly, he simply retreats to ideal theory. Scientists are not only concerned with what is epistemically preferable, but what is preferable all things considered. To suggest otherwise is both descriptively inadequate and normatively question-begging. We don’t want an epistemic ideal, but a *scientific* ideal, that is, an ideal to guide scientists who have both epistemic and social responsibilities. By retreating from arguments about what is all-things-considered preferable, Hudson does not successfully offer an argument that scientists should be guided by the VFI. (Alternatively, Hudson may be gesturing at a “pursuitworthiness” defense of the VFI; see next section.)

Sheykh-Rezaee and Bikaraan-Behesht (2023) also argue that the VFI is an “epistemic ideal.” Their argument amounts to a kind of burden-shifting argument. Rather than attack the premises of the AIR, with which they identify no fault, they argue that, despite the conclusion that “non-epistemic values have a legitimate role to play in the internal stages of science,” (even a *necessary* role, they concede), this is not sufficient to defeat the VFI, because it does not meet their “constitutivity” condition: “The role the non-epistemic values necessarily play at the internal stages of science should be shown to be *constitutive*” (146). Although non-epistemic values play a legitimate, even *necessary* role in scientific reasoning, they do not play a *constitutive* role, and so the VFI is supposedly unscathed. It is not particularly clear what constitutivity amounts to, nor is it well-motivated as a criterion. More importantly, it has no relevance to the kinds of decisions scientists make in practice, because they concede that it may be all-things-considered necessary for scientists to consider non-epistemic values in the internal stages of science. An ideal for scientists must be able to *guide* scientific practice. Labeling parts of scientific practice “non-constitutive” is idle.

6.4.4 Pursuitworthiness of the VFI

Menon and Stegenga (2023; Stegenga and Menon 2023) provide a novel defense of the VFI based on its pursuitworthiness; they argue that the VFI might be worth pursuing, even if unattainable. That is, unattainability does not imply that it is not pursuitworthy. This is widely acknowledged in the literature (based on the definition of “ideals”), though opponents of the VFI obviously argue that it still does not turn out to be good to pursue the VFI. Menon and Stegenga also argue, intriguingly, that the VFI might be worth pursuing even if it would be undesirable *in principle* to *achieve* it. They cast the VFI as aiming at the elimination of “bifurcation points,” i.e., decision points where differences in value judgments made in the course of inquiry would lead scientists to infer different conclusions from the same evidence. On the other hand, they acknowledge that it is legitimate for values to play a role in many upstream decisions that determine both how conclusions are framed and what evidence is available. According to their interpretation of the VFI, values should not make a “difference to inference” (Stegenga and Menon 2023).

Menon and Stegenga do not argue that science can or *should* be value-free. They acknowledge that the AIR shows that values should play a role in science in some cases. It is tempting to classify this as a kind of partial rapprochement with the VFI rather than a defense per se, but they insist that it is instead a legitimate alternative framing of the VFI, according to which: “Scientists should [typically] act *as if* science should be value-free” (Menon and Stegenga 2023, *emphasis added*; they call this version of the VFI “VFI4”). Between their reframing of the ideal and of the stakes of the debate,

they hope to obviate the key arguments against VFI, including those canvassed by [Elliott \(2026\)](#). One tempting response here is that they have simply given up the game on the VFI. In explicitly denying the need to rebut the AIR, they concede the main issue. This is too hasty, however, as there is a substantive issue when it comes to providing action-guiding advice to scientists: should scientists make (appropriately warranted or constrained, non-epistemic) value judgments as a routine part of their inquiries, or should they behave as if science ought to be value-free, and thus typically deploy strategies that avoid value judgments?

Menon and Stegenga provide three reasons to think that value-freedom is worth pursuing even if unattainable and undesirable as an end-state: (i) when values influence scientific reasoning, it becomes less truth-apt; (ii) value-free science advising is more democratic; (iii) public trust in science depends on value-freedom. Their argument for (i) involves cherry-picking examples of bad value-laden science while denying that positive examples are probative. It also wrongly assumes, without argument, that ethical and political values are not truth-apt; (iii) is not supported by empirical studies on the relation of trust and value-ladenness ([Hicks and Lobato 2022](#)); (ii) and (iii) are also versions of the democratic defense of VFI, which, I have argued, does not provide a compelling defense of the VFI so much as a desideratum for answers to the value-management question; in this case, the advice is to minimize their influence unless doing so runs afoul of more important constraints, such as those, “concerning resource use, research ethics and action-guidance” ([Menon and Stegenga 2023](#)).

When it comes to practical advice for scientists, Menon and Stegenga suggest that scientists should typically adopt “value-mitigating strategies,” i.e., strategies that will eliminate bifurcation points, except in cases where moral and practical constraints make adopting those strategies undesirable. The key problem with this argument is that we cannot tell ahead of time which case we are in; so, in fact, scientists will have to continue weighing non-epistemic values throughout inquiry in order to determine whether value-mitigation is permissible or preferable over explicit value judgment. This undermines the idea that this approach is value-free (even in their *as-if* sense). [Stegenga and Menon \(2023\)](#) pursue a similar strategy, but add the argument that values in science impede the achievement of scientific consensus. The argument is problematic in at least two respects. First, it is not at all clear that science ought to strive for consensus, as countless pluralists have argued. Second, there is no necessary connection between values and consensus. Values might play exactly the role the AIR specifies, but not interfere with strong consensus ([Stegenga and Menon 2023](#), 435; cf. [Kourany 2010](#)). While it is easy to be pessimistic in the face of disputes over values, hope for reaching sufficient agreement is at least as attractive a regulative ideal as the VFI. What’s more, potentially any consideration that factors in inference decisions leads to dissensus; values are no more likely than different hypotheses, data analysis techniques, sets of evidence, or epistemic criteria to frustrate consensus. It is best to think about what Menon and Stegenga are doing as answering the value-management question, and judge its merits in comparison to other answers to that question.

6.5 Partial rapprochements with the VFI

There is a set of arguments that, while not exactly defending the VFI, seek to accommodate what they take to be right in both defenses and critiques of the VFI by specifying value-laden and value-free moments of scientific inquiry, while acknowledging that the result is thus value-laden in a sense. Despite somewhat different aims, defenders of such views imply that outright rejection of the VFI goes too far, seeking to preserve some degree or form of value-freedom while acknowledging the force of critiques of the VFI. However, this rhetoric typically presupposes a kind of error, the conflation of the VFI question with the value-management question. One can take a quite conservative view on the value-management question (e.g., [Steel 2010](#)), but this does not uphold the VFI. (Note that the error works both ways; [de Melo-Martín and Intemann \(2016\)](#) make this mistake when they argue that the

AIR does not go far enough, and so Douglas's positive account *vindicates* the VFI. Their disagreement is actually with Douglas's response to the value-management question, which presupposes a negative answer to the VFI question.)

Some attempted rapprochements focus on sequestering value influences into a certain phase of scientific inquiry. Recent work by Kareem Khalifa and collaborators, for instance, focuses on the role of questions in inquiry, and explicitly aligns itself with the value-free ideal in the sense of taking decisions to give or accept an answer to a question to only be rationally determined by "epistemic considerations" (DiMarco and Khalifa 2019, 1022; cf. Khalifa, Millson, and Risjord 2025). In other words, they reconstruct a version of the context of discovery/context of justification distinction into the context of posing questions and the context of answering questions, arguing in effect that, while one may not be able to restrict all influence of values to the context of discovery, one *can* so limit it to the context of posing questions.

There are several problems with this proposal *as* a rapprochement with the VFI. First, it is not always possible to specify in advance the range of value considerations, so that they can be loaded into the question that inquiry begins with. The research question that inquiry is trying to solve may be only vague and inchoate at the start of inquiry, and it co-evolves with the tentative answers, where neither is properly settled until inquiry concludes (Brown 2020, Ch. 1). The relevant options and values may likewise only be discovered in the course of inquiry, not settled ahead of time. Khalifa et al. can acknowledge this point but insist that they can retain their view by packing the values into the posing of new questions *throughout* the course of inquiry. There remain several further concerns. One is Okruhlik's "values in, values out" problem; Okruhlik (1994), in a very important essay, argues that any value-free decision procedure for comparing hypotheses or theories, operating on inputs that are value-laden, will reproduce rather than eliminate that value-ladenness (cf. Elliott and McKaughan 2009). Perhaps most pressingly, it seems that this approach is merely a notational or terminological variant of the picture presented by the opponents of the VFI, describing in different terms how value judgments figure in the internal processes of scientific inquiry, rather than denying that they do. This is not to say that such alternative analyses have no value, only that they do not in fact partly vindicate the VFI; rather, they address the value-management question in a particular way.² Furthermore, assessment of answers to questions are themselves liable to the same kinds of concerns about the regress of upstream and downstream inductive risks as are the assessments of probabilities assigned to hypotheses given evidence as advocated by the hedging response.

A related approach focuses on the concept of adequacy-for-purpose in model assessment, which supposedly "opens the possibility of effecting a partial rapprochement between critics and proponents of the value-free ideal" (Lusk and Elliott 2022). Applying the adequacy-for-purpose approach from model assessment more widely to all forms of scientific assessment, Lusk and Elliott argue that we may be able to recast what looks like value-laden scientific reasoning into value-free assessment of hypotheses about whether something is adequate for some purpose, though, as they note, such assessments (may) remain susceptible to inductive risk concerns. Yet another approach of this type, based on adequacy-for-purpose judgments, is Wendy Parker's (2024) recent "epistemic projection approach." These approaches are susceptible to many of the same concerns as Khalifa et al., applying the arguments concerning loading values into "questions" with loading them into "purposes." By contrast, Harvard and Winsberg (2022) correctly note that an adequacy-for-purpose view means that model assessment is ineliminably value-laden.

These attempts at partial rapprochement with the VFI are nonetheless generally valuable contributions; it would just be better to frame them as answering the value-management question. It is not that we need to find a middle ground in the debate for and against the VFI, because the VFI is rightly understood as an "ideal of epistemic purity" (Biddle 2013), and so any "middle ground" *is* a rejection

of the VFI. Rather, they raise important considerations when addressing the value-management question, and in many cases suggest serious frameworks for answering that question (if not entirely unproblematic ones). The urge toward rapprochement should be rejected in favor of more careful analysis of the issue of managing the roles of values in science.

6.6 Conclusions

For all the current interest behind them, arguments in defense of the VFI either fail entirely or prove not to be best understood as arguments precisely against the VFI. What is live in these arguments supports the general consensus in the field that the VFI is dead, and the interesting questions concern how best to manage the role of values in science.

Acknowledgments

This chapter is adapted from a longer paper, “For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science,” published in *Synthese* (Brown 2024). My gratitude to the editors of this volume for their flexibility and generosity in bringing my contribution to this volume to fruition.

Notes

- 1 My gratitude to Jacob Stegenga for drawing my attention to this trade-off in a recent presentation.
- 2 Indeed, Khalifa in particular has said that he does not intend this work to side with the VFI against the AIR (Personal communication).

References

- Alexandrova, Anna, and Mark Fabian. 2022. “Democratising Measurement: Or Why Thick Concepts Call for Coproduction.” *European Journal for Philosophy of Science* 12(1): 7. <https://doi.org/10.1007/s13194-021-00437-7>.
- Betz, Gregor. 2013. “In Defence of the Value Free Ideal.” *European Journal for Philosophy of Science* 3 (2): 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Biddle, Justin B. 2013. “State of the Field: Transient Underdetermination and Values in Science.” *Studies in History and Philosophy of Science* 44 (1): 124–33. <http://dx.doi.org/10.1016/j.shpsa.2012.09.003>.
- Boulcault, Marion, and S. Andrew Schroeder. 2021. “Public Trust in Science: Exploring the Idiosyncrasy-Free Ideal.” In *Social Trust*, edited by Kevin Vallier and Michael Weber. Routledge.
- Bright, Liam Kofi. 2018. “Du Bois’ Democratic Defence of the Value Free Ideal.” *Synthese* 195 (5): 2227–45. <https://doi.org/10.1007/s11229-017-1333-z>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Brown, Matthew J. 2024. “For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science.” *Synthese* 204 (112). <https://doi.org/10.1007/s11229-024-04762-1>.
- Brown, Matthew J., and Jacob Stegenga. 2023. “The Validity of the Argument from Inductive Risk.” *Canadian Journal of Philosophy* 53 (2): 187–90. <https://doi.org/10.1017/can.2023.37>.
- Carrier, Martin. 2022. “What Does Good Science-Based Advice to Politics Look Like?” *Journal for General Philosophy of Science* 53 (1): 5–21. <https://doi.org/10.1007/s10838-021-09574-2>.
- Cassini, Alejandro. 2022. “Simulation Models and Probabilities: A Bayesian Defense of the Value-Free Ideal.” *Simulation* 98 (2): 113–25. <https://doi.org/10.1177/00375497211028815>.
- de Melo-Martín, Inmaculada, and Kristen Intemann. 2016. “The Risk of Using Inductive Risk to Challenge the Value-Free Ideal.” *Philosophy of Science* 83 (4): 500–20. <https://doi.org/10.1086/687259>.
- DiMarco, Marina, and Kareem Khalifa. 2019. “Inquiry Tickets: Values, Pursuit, and Underdetermination.” *Philosophy of Science* 86 (5): 1016–28. <https://doi.org/10.1086/705446>.
- Douglas, Heather. 2000. “Inductive Risk and Values in Science.” *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.

- Douglas, Heather. 2005. "Inserting the Public into Science." In *Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making*, edited by Sabine Maasen and Peter Weingart. Sociology of the Sciences Yearbook. Springer.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2012. "Weighing Complex Evidence in a Democratic Society." *Kennedy Institute of Ethics Journal* 22 (2): 139–62. <https://doi.org/10.1353/ken.2012.0009>.
- Douglas, Heather. 2016. "Values in Science." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.
- Edenhofer, Ottmar, and Martin Kowarsch. 2015. "Cartography of Pathways: A New Model for Environmental Policy Assessments." *Environmental Science & Policy* 51: 56–64. <https://doi.org/10.1016/j.envsci.2015.03.017>.
- Elliott, Kevin C. 2011. *Is a Little Pollution Good for You?: Incorporating Societal Values in Environmental Research*. Oxford University Press.
- Elliott, Kevin C. 2026. "Arguments against the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. "How Values in Scientific Discovery and Pursuit Alter Theory Appraisal." *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Fernández Pinto, Manuela, and Daniel J. Hicks. 2019. "Legitimizing Values in Regulatory Science." *Environmental Health Perspectives* 127 (3): 035001. <https://pmc.ncbi.nlm.nih.gov/articles/PMC6768319/>.
- Frisch, Mathias. 2020. "Uncertainties, Values, and Climate Targets." *Philosophy of Science* 87 (5): 979–90. <https://doi.org/10.1086/710538>.
- Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32 (1): 1–31. <https://doi.org/10.1353/ken.2022.0001>.
- Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Havstad, Joyce C., and Matthew J. Brown. 2017. "Inductive Risk, Deferred Decisions, and Climate Science Advising." In *Exploring Inductive Risk*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Henschen, Tobias. 2021. "How Strong Is the Argument from Inductive Risk?" *European Journal for Philosophy of Science* 11 (3): 92. <https://doi.org/10.1007/s13194-021-00409-x>.
- Hicks, Daniel J., and Emilio Jon Christopher Lobato. 2022. "Values Disclosures and Trust in Science: A Replication Study." *Frontiers in Communication* 7. <https://doi.org/10.3389/fcomm.2022.1017362>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Hudson, Robert. 2016. "Why We Should Not Reject the Value-Free Ideal of Science." *Perspectives on Science* 24 (2): 167–91. https://doi.org/10.1162/posc_a_00199.
- Jeffrey, Richard C. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science* 23 (3): 237–46. <https://doi.org/10.1086/287489>.
- Kappel, Klemens. 2014. "The Proper Role of Science in Liberal Democracy." *Unpublished Draft*. Accessed June 3, 2025 at <https://www.academia.edu/7017103/>.
- Kappel, Klemens, and Julie Zahle. 2019. "The Epistemic Role of Science and Expertise in Liberal Democracy." In *The Routledge Handbook of Social Epistemology*, edited by Miranda Fricker, Peter J. Graham, David Henderson, and Nikolaj J. L. L. Pedersen. Routledge.
- Khalifa, Kareem, Jared Millson, and Mark Risjord. 2025. "Inquiry and Epistemic Priority." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw, and Carla Fehr. Boston Studies in the Philosophy and History of Science. Springer.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Kourany, Janet A. 2010. *Philosophy of Science After Feminism*. Oxford University Press.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Lusk, Greg, and Kevin C. Elliott. 2022. "Non-Epistemic Values and Scientific Assessment: An Adequacy-for-Purpose View." *European Journal for Philosophy of Science* 12 (2): 35. <https://doi.org/10.1007/s13194-022-00458-w>.
- MacGillivray, Brian H. 2019. "Null Hypothesis Testing Scientific Inference: A Critique of the Shaky Premise at the Heart of the Science and Values Debate, and a Defense of Value-Neutral Risk Assessment." *Risk Analysis* 39 (7): 1520–32. <https://doi.org/10.1111/risa.13284>.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom Is Worth Pursuing." *European Journal for Philosophy of Science* 13 (4): 48. <https://doi.org/10.1007/s13194-023-00552-7>.

- Mitchell, Sandra. 2004. "The Prescribed and Proscribed Values in Science Policy." In *Science, Values, and Objectivity*, edited by Peter Machamer and Gereon Wolters. University of Pittsburgh Press.
- Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24 (S1): 21–42.
- Parker, Wendy S. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36. <https://doi.org/10.1017/psa.2023.107>.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/28723>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Sheykh-Rezaee, Hossein, and Hamed Bikaraan-Behesht. 2023. "Value-Free Ideal Is an Epistemic Ideal: An Objection to the Argument from Inductive Risk." *Principia: An International Journal of Epistemology* 27 (1): 137–63. <https://doi.org/10.5007/1808-1711.2023.e80487>.
- Silk, Matthew. 2018. "Evaluation and Value Management in Science." PhD diss., University of Waterloo. <https://dspacemainprd01.lib.uwaterloo.ca/server/api/core/bitstreams/cf43560a-00c5-423d-9fe2-c111bfff3193d/content>.
- Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. <https://doi.org/10.1086/650206>.
- Steel, Daniel. 2015. "Acceptance, Values, and Probability." *Studies in History and Philosophy of Science Part A* 53: 81–88. <https://doi.org/10.1016/j.shpsa.2015.05.010>.
- Steel, Daniel. 2016. "Climate Change and Second-Order Uncertainty: Defending a Generalized, Normative, and Structural Argument from Inductive Risk." *Perspectives on Science* 24 (6): 696–721. https://doi.org/10.1162/POSC_a_00229.
- Steele, Katie. 2012. "The Scientist Qua Policy Advisor Makes Value Judgments." *Philosophy of Science* 79 (5): 893–904. <https://doi.org/10.1086/667842>.
- Stegenga, Jacob, and Tarun Menon. 2023. "The Difference-to-Inference Model for Values in Science." *Res Philosophica* 100 (4): 423–47. <https://doi.org/10.5840/resphilosophica2023928102>.
- Ward, Zina B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science Part A* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Wilholt, Torsten. 2023. "Harmful Research and the Paradox of Credibility." *International Studies in the Philosophy of Science* 36 (3): 193–209. <https://doi.org/10.1080/02698595.2023.2273606>.
- Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–37. <https://doi.org/10.1353/ken.2012.0008>.
- Winsberg, Eric. 2024. "Managing Values in Science: A Return to Decision Theory." *Kennedy Institute of Ethics Journal* 34(4): 389–418. <https://doi.org/10.1353/ken.2024.a965814>.

7

WHAT DOES IT MEAN TO SAY THAT SCIENCE IS VALUE-LADEN?

Zina B. Ward

7.1 Introduction

The claim that science involves non-epistemic values is, in one sense, fairly trivial. Science is an activity we pursue for the sake of both epistemic and non-epistemic ends. We care about producing knowledge or understanding, but we also want to promote human health, develop commercially lucrative products, and achieve public policy goals. Decisions about which scientific projects to pursue—what research is deserving of our finite resources—are shaped by non-epistemic judgments of practical importance. We devote more money to studying diseases that affect humans than diseases that affect koalas, and rightly so. Science is also subject to constraints from non-epistemic values, such as in the design of research, where requirements to safeguard the well-being of human or animal subjects can trump the demand for new knowledge.

Locating non-trivial claims about the role of values in science requires a finer grain of analysis. Rather than considering science as a whole, interesting philosophical questions arise when we ask which scientific choices involve non-epistemic values and how such choices should be handled. This chapter focuses on claims of the former sort, which state that a particular type of scientific choice is value-laden (or not).¹ Over the decades, philosophers have put forward many such claims. Recently, for example, [Elabbar \(2023\)](#) has argued that the curation of evidence in scientific assessments is value-laden, and [Dethier and Fletcher \(2024\)](#) have contended that the selection of an estimator in classical statistical inference is value-laden.

There is often a hidden complexity in claims of this sort, arising from ambiguity in what it means for a choice to be “value-laden.” My aim in this chapter is to provide a disambiguation of value-ladenness and offer a novel demonstration of its usefulness. In [Sections 7.2 and 7.3](#), I distinguish four ways a choice can be said to be value-laden. [Section 7.4](#) discusses the relationship between this taxonomy and recent work on the ontology of values. In [Sections 7.5 and 7.6](#), I show that the taxonomy provides tools for a nuanced assessment of the “randomizer reply” to arguments for the inevitability of value-laden choices in science.

7.2 The space of reasons and the space of causes

In his essay “Empiricism and the Philosophy of Mind,” Wilfrid Sellars introduces a now well-known philosophical dichotomy. “In characterizing an episode or state as one of knowing,” he writes, “we are not giving an empirical description of that episode or state: we are playing in the logical space of reasons,

of justifying and being able to justify what one says” (Sellars 1956, 298–9). Later commentators, including Rorty (1979), McDowell (1996), and Brandom (1998), elaborated on Sellars’ notion of the space of reasons as the arena in which our actions or attitudes are normatively evaluated. The space of reasons is distinct from the space of causes, which is delivered by the sciences and contains descriptions and explanations of worldly phenomena. Understanding the relationship between the space of reasons and the space of causes is a central project for philosophers interested in naturalism and normativity (Pollard 2005).

The notion of “value” at play in work on values in science is notoriously slippery. McMullin (1982), for instance, writes:

“Value” is one of those weasel words that slip in and out of the nets of the philosopher. We shall have to try to catch it first, or else what we have to say about the role of values in science may be of small use.

Many definitions of “value” have been offered, but none have been entirely satisfying or widely adopted. I agree with McMullin that this is a problem. How can we determine the (proper) role of values in science if we do not have a grasp of values themselves? In Ward (2021), I aimed to put the literature on a firmer footing by analyzing how values bear on choices. The first step of this analysis can be put in Sellarsian terms.

In some ways of thinking about values, they play a role in the space of reasons. Values often justify our actions—as when, for example, the value of pleasure justifies my choice of morning coffee. In other ways of thinking about values, they play a role in the space of causes. We talk about values causing things—as in your unconscious value for attracting attention causing you to buy a particular car. The notion of a value thus seems to admit of different interpretations in different contexts. Philosophers of science have overlooked this variability to their peril, leading to crosstalk and misunderstanding.

7.3 Four senses of value-ladenness

Values can either act as reasons for choice or interact causally with choices. Within each of these categories, a further distinction can be made, yielding a four-way taxonomy:

Values in the space of reasons

- 1 *Rational interpretation* (values act as justifying reasons):

A choice is value-laden when it is justified by non-epistemic values.

- 2 *Motivational interpretation* (values act as motivating reasons):

A choice is value-laden when it is motivated by non-epistemic values.

Values in the space of causes

- 3 *Causal interpretation* (values act as causal effectors):

A choice is value-laden when it is causally influenced by the possession of certain non-epistemic values.

- 4 *Objectual interpretation* (values act as affected goods):

A choice is value-laden when it causally promotes or impedes a non-epistemic value.

Let's unpack each interpretation with a few examples. Philosophers of action distinguish between two sorts of reasons: motivating reasons and justifying reasons (Baier 1958; Bond 1974). The former motivate a person while the latter justify what they do. The two sorts of reasons are not always the same. For instance, imagine that Peter decides to become a vegetarian. There are plausibly several reasons that justify Peter's decision: cutting meat out of his diet reduces his carbon footprint, prevents animal suffering, and helps him fit in with his vegetarian friends. As it happens, only the last reason actually motivates Peter; it is because of his desire to fit in that he becomes a vegetarian. Some of the justifying reasons for his choice are thus not its motivating reason.

Non-epistemic values often act as motivating and/or justifying reasons. Indeed, the same example can also be characterized using the language of values: Peter's choice is motivated by a value for social acceptance, although it is also justified by values for the environment and animal welfare. There are two corresponding ways in which a choice can be said to be value-laden. The first is when non-epistemic values act as justifying reasons for the choice. The second is when non-epistemic values (consciously or unconsciously) act as motivating reasons for the choice. These are the *rational* and *motivational interpretations* of value-ladenness.

Decisions about which scientific research projects to fund are arguably value-laden in a rational sense, since they are justified, at least in part, by non-epistemic values. A project that aims to estimate the number of ants in my backyard and a project that aims to understand which genetic mutations affect the shape of the COVID-19 spike protein would both produce new knowledge. It's not because of an epistemic difference between the projects, but rather a practical and moral difference, that the decision to fund the latter but not the former is justified (Kitcher 2001). Note that claims about rational value-ladenness require there to actually be a justificatory relationship between the value and choice in question. The decision to fund research on the COVID-19 spike protein is rationally value-laden because the value of human health does indeed justify undertaking such research (though the justification is only *pro tanto*; see Ward 2021).

Meanwhile, many scientists' individual decisions about what research to pursue are value-laden in the motivational sense. Consider a parasitologist who decides to specialize in the fungus *Cryphonectria parasitica*, commonly known as chestnut blight, because of its devastating impact on her historic hometown and her hope that scientific advances can protect against future invasive pathogens. The parasitologist's choice of specialization is value-laden in that ecological and social values serve as motivating reasons for her. (As is true for each example in this section, the choice might be value-laden in other senses as well.)

In addition to playing these roles in the space of reasons, values can also operate in the space of causes. Here too there is a natural bifurcation: values can be understood to play the role of either cause or effect. That is, the possession of a value can affect a choice, or a choice can have an impact on goods in the world. On the first, *causal interpretation* of value-ladenness, a choice is value-laden when someone's (or some group's) possession of a particular non-epistemic value influences the choice. Imagine that Keira's family is getting a dog. Wanting to economize, Keira selects the ten dogs that would be cheapest to adopt from the local animal shelter. She gives this list of ten dogs to her son Randall, who looks at their photos and chooses the cutest. Randall's choice of dog is influenced by Keira's possession of the value of frugality, and so is causally value-laden, even though frugality is not among his motivations. Frugality plays a causal but not a motivational role with respect to Randall's choice.

There are countless scientific choices that are analogously value-laden in a causal sense. In 2015, decision-makers at the National Institutes of Health (NIH), concerned about the historical exclusion of female organisms from biomedical research, adopted a policy encouraging researchers to collect data on and disaggregate their data by sex. This "sex as a biological variable" guidance has shaped a huge number of scientific choices in the years since—choices about project selection, data collection, statistical analysis, and presentation of findings. Such choices, impacted as they are by the NIH's

possession of a value for gender equality, are value-laden in a causal sense. This is so even when scientists are not themselves motivated by a concern for equality or even aware of the values that shaped the guidelines they are following.

On the final, *objectual interpretation* of value-ladenness, a choice is value-laden if it has an impact on non-epistemic goods in the world—goods that we sometimes call “values.” Imagine that a group of city commissioners decides to have pickleball courts built in a city park. As it turns out, the building of the courts inflames old rivalries. The city’s factions start fighting bitterly about line calls and court time, disputes that bleed into other areas of life. In this scenario, the decision to build the pickleball courts is objectually value-laden because it (negatively) affects the value of community cohesion, a non-epistemic good.

Since our choices often have non-epistemic impacts, objectually value-laden choices are common in everyday life and in science. Consider a team of meteorologists who develop a complex weather model to predict storm surge along a densely populated coastline. Their model becomes the basis for many local governments’ evacuation orders. Choices during model development that influence its outputs, and thereby local governments’ policies, are value-laden in the objectual sense: they affect non-epistemic goods in the world, such as the preservation of human life.

I make no claim to the exhaustiveness of this taxonomy, nor the mutual exclusivity of its four categories. A choice can easily be value-laden in multiple senses at once. For instance, it is plausible that the city commissioners’ decision to build pickleball courts is motivationally as well as objectually value-laden; the parasitologist’s choice to study *C. parasitica* is both rationally and motivationally value-laden; and Peter’s becoming a vegetarian is value-laden in all four senses.

The taxonomy also leaves open a number of contested questions in action theory and moral philosophy. For instance, philosophers of action disagree about whether a person’s motivating reasons are causes of their actions (Davidson 1963). If they are, then motivational claims about value-ladenness are also causal claims about value-ladenness. Some would argue, for example, that because the parasitologist’s choice is motivated by non-epistemic values, it is causally value-laden as well. (As the examples above show, the reverse does not hold: someone’s possession of a value can causally influence a downstream choice without being a motivating reason for that choice.)

A related question is whether values that act as motivating reasons really operate in the space of reasons. If one believes that motivating reasons are causes, then motivational claims about value-ladenness concern the space of causes. Here too I wish to remain agnostic, as I am more committed to the taxonomy’s four lower-level categories than its overarching bipartite division. Depending on one’s other commitments, one may take there to be just one sense of value-ladenness associated with the space of reasons (the rational interpretation).

7.4 Value-ladenness and the ontology of values

My taxonomy is noncommittal about the ontology of values, focusing instead on how values bear on choices. Ward (2021) did float a more “metaphysically ambitious” view, on which “there are not only several different relationships between values and choices, but several distinct kinds of things we call values... [V]alues are not one sort of entity, but four” (57). Such an account mirrors the view in philosophy of action that, instead of reasons merely playing two different “roles,” motivating reasons and justifying reasons are really two “quite different sorts of thing” (Bond 1974, 333). Perhaps values, too, are four quite different sorts of thing. Instead of characterizing values merely in terms of the roles they play, one might claim that values *just are* justifying values, motivating values, causal effectors, or affected goods.

The ontology of values has recently received direct attention from philosophers working on values in science (Biddle 2013; Brown 2020; Hilligardt 2022). In “Values in Science: What Are Values, Anyway?”, Elliott and Korf (2024) distinguish “four ways of conceptualizing values” (4). First, they

argue, some authors follow Kuhn (1977) in conceiving of values as “criteria or standards for choice.” Kuhn was concerned with criteria of theory choice, but others think of values more expansively as “criteria for evaluating a broader array of entities” (Elliott and Korf 2024, 4–5). Second, values can be thought of as “factors that causally influence scientists’ choices, habits, and practices” (Elliott and Korf 2024, 5). They remain agnostic about whether such factors are all psychological. Third, values are sometimes taken to be “beliefs or attitudes about what is desirable” (Elliott and Korf 2024, 6). They suggest that there is overlap between the second and third conceptions of values: some causal factors are beliefs or attitudes about what is desirable. Fourth, values can be seen as “desirable things themselves” (Elliott and Korf 2024, 6). Like Ward (2021), Elliott and Korf argue that, because different authors operate with different value concepts, the literature on values in science is less unified than is commonly thought.

How does Elliott and Korf’s taxonomy compare to mine? As I see it, there is a close correspondence between our categories. Recall that the causal interpretation of value-ladenness concerns values in their role as causal effectors, whose possession influences scientific choices. The objectual interpretation of value-ladenness concerns values conceived of as non-epistemic goods in the world, which can be promoted or impeded. Elliott and Korf’s concepts of values as causal factors and desirable things straightforwardly map onto these two interpretations.

Their concept of values as beliefs or attitudes about desirability corresponds in a looser way to the motivational interpretation, which treats values as motivating reasons. The reasons that motivate our actions are often grounded in our beliefs or attitudes about what is desirable. Fourth and finally, the rational interpretation of value-ladenness takes values to act as justifying reasons, “reasons for or against doing a thing” which are “tied to the world beyond” rather than to a person’s motivations (Bond 1983, 30). Elliott and Korf’s concept of values as criteria or standards for choice, which they also call “*reasons for choice*,” evokes this rational interpretation (2024, 6; my italics). In addition to emphasizing our shared language of “reasons,” I suggest that criteria or standards for choice are things “tied to the world” in Bond’s sense, rather than to an individual’s psychology. Standards and criteria justify or rationalize choices, and this, of course, is exactly what justifying reasons do.

Given these mappings between the categories in our taxonomies, I take Elliott and Korf to be identifying the typical *realizers* of my four *roles* for values. That is, they are pointing to the entities that often play each of the roles distinguished here and in Ward (2021). For example, those things which fulfill the role of justifying choice are often standards and criteria; what fulfills the role of causally influencing choices are (somewhat trivially) causal factors; and so on. Elliott and Korf, too, note some of the similarities between our taxonomies but argue that the concept of value that one deploys is dissociable from what role the value plays. For example, a belief or attitude about what is desirable could serve as a causal effector, justifying reason, or motivating reason.² Evaluating such claims would take us too deep into debates about the ontology of reasons. (Do beliefs and attitudes justify? Can criteria or standards be causes, or do they belong solely to the space of reasons?) But note that their claim is compatible with the view that their four “concepts” are at least the typical realizers of values’ four roles.

7.5 Inevitability arguments and the randomizer reply

Philosophers’ failure to distinguish different senses of value-ladenness has led to spurious disagreements. One author will argue that a particular scientific choice is value-laden, and another will deny that it is value-laden. But their disagreement is merely apparent because they mean different things by “value-laden.” Since scientific choices can be value-laden in one sense but not another, both authors can be right. For example, a choice may be value-laden in the objectual sense but not the motivational sense: plenty of scientific choices affect goods in the world even though the people making them are

not motivated by non-epistemic considerations. Other scientific choices are value-laden in the causal sense but not the motivational sense: someone's possession of non-epistemic values has a causal effect on a later decision without the decision-maker herself being motivated by those values. Many debates about values in science could therefore be clarified or dissolved by paying closer attention to the notions of value-ladenness at play.

I'll close this chapter with an extended illustration involving a family of what I'll call "inevitability arguments." Inevitability arguments assert that a certain scientific choice *must be* value-laden (in a non-moral sense), i.e., that the choice inevitably implicates non-epistemic values. Proponents of inevitability arguments begin by identifying a scientific choice that is not fully determined by evidence or epistemic values alone: there is a gap between what epistemic considerations dictate and what a scientist must eventually decide.³ They then claim that non-epistemic values are inevitably involved when scientists face such "unforced" choices (Winsberg 2012). Since epistemic values underdetermine the choice, non-epistemic values are necessarily implicated in selecting among the epistemically acceptable options.

The conclusion of inevitability arguments is quite strong: they insist that values *have to be* involved in the scientific choice under discussion (not that they *may* or *should be* involved). At the same time, inevitability arguments encompass a variety of ways of arguing for the value-ladenness of science. The following are all inevitability arguments:

- versions of the argument from inductive risk that claim that non-epistemic values have to be involved in setting the threshold for when the evidence is sufficient to accept a hypothesis (Biddle and Kukla 2017; Rudner 1953)
- feminist underdetermination arguments which hold that, because there are multiple background assumptions one could use to assess evidential relevance, non-epistemic values must be invoked to choose between them (Longino 1990)
- arguments which claim that, because epistemic values must be interpreted and traded off against one another, the deployment of epistemic values has to involve non-epistemic values (Hoynigen-Huene 1992; Rooney 1992; cf. Ward 2024)

Each of these arguments points to a particular type of scientific choice—about evidential thresholds, background assumptions, or epistemic values—that it claims is inevitably value-laden because epistemic considerations underdetermine the choice.

One way to resist inevitability arguments is to deploy what Magnus (2018) calls "the randomizer reply." The reply points out that, when faced with an unforced choice, a scientist has the option of selecting randomly among the epistemically acceptable options—say, by rolling a die or flipping a coin. Since there is an alternative to using non-epistemic values to break the underdetermination, inevitability arguments fail to establish that scientific choices are necessarily value-laden. (Note that the randomizer reply does not claim that scientific decisions *should* be made randomly. It only insists that random decision-making is a theoretical possibility, thereby undermining inevitability arguments. One could accept the randomizer reply's contention that unforced scientific choices *need not* be value-laden, while also believing that, as a practical matter, such choices *should be* made by appeal to values.)

Proponents of the randomizer reply include Betz (2013), who rejects the claim that "Scientists inevitably make non-epistemic value judgments when establishing (adopting and communicating) policy-relevant results" (210). A defender of the value-free ideal, Betz argues as follows: "[S]uppose that the scientist, when facing an arbitrary choice, simply rolls a die. It's at least not straightforward which specific, non-epistemic normative assumptions she (implicitly) buys into by doing so" (Betz 2013, 210). He treats this possibility as a decisive objection. de Melo-Martín and Intemann (2016) note that there is ambiguity in the claim that, as some versions of the argument from inductive risk

would have it, non-epistemic values are “necessary” for assessing the consequences of error. One interpretation of this claim invokes logical necessity: it is “logically impossible to derive conclusions about what hypotheses to accept or reject without relying on value judgments” (2016, 505). de Melo-Martín and Intemann quickly reject this as implausible, “as there are other logically possible ways to arrive at conclusions, such as flipping a coin” (2016, 505).⁴

Others have rejected the randomizer reply (Douglas 2021; Magnus 2018; Havstad 2022). They argue that, instead of avoiding the inevitable involvement of non-epistemic values in science, randomization simply relocates it. This is because the decision to deploy randomization is itself necessarily driven by values. Here is Magnus (2018):

[C]onsider a mundane case in which I cannot decide which of two restaurants to visit for lunch and so flip a coin. My values and practical reasons have no influence over the outcome of the coin toss, of course, and so the selection is value-free to that extent. However, my values and preferences are involved in my decision to use coin-flipping as a way of resolving the choice. I want to go to lunch at one of two places, and I do not want to spend too much time or energy deciding. If someone asks why I went to one restaurant rather than the other, a complete answer would refer not just to the random process but also to the reasons I had for adopting that method. Similarly, deciding to flip a coin in the face of underdetermination would be practical and value-driven. Breaking ties by flipping coins would keep values from directly deciding specific winning hypotheses, but it would not ultimately escape the intrusion of values and practical decisions into theory choice. (416)

This, then, is the current state of play: proponents of inevitability arguments claim that certain scientific choices are inevitably value-laden; advocates of the randomizer reply point out that such choices can in principle be made randomly, without involving values; and defenders of inevitability arguments offer a rejoinder, rejecting the idea that randomization circumvents value-ladenness.

7.6 A taxonomy-guided assessment

Each move in this debate can be fruitfully assessed by attending to what different authors mean by “value-laden.” It turns out that the plausibility of inevitability arguments and the randomizer reply depend on which conception of value-ladenness one adopts.

First, consider an objectual interpretation, under which inevitability arguments claim that how one makes an unforced scientific choice has an inevitable effect on non-epistemic goods in the world. This is surely true of at least some choices in consequential areas of science. For such choices, the randomizer reply can be quickly dismissed. If a scientific decision will inevitably have a non-epistemic impact, it makes no difference whether that decision is made randomly. Randomization is thus a dead-end when it comes to rebutting objectual claims about the unavoidable role of values in science.

The randomizer reply is more convincing if inevitability arguments are interpreted in motivational terms. Such arguments claim that, because epistemic considerations fail to dictate a scientist’s choice, she must be motivated to select the option she does by non-epistemic values. This argument is undermined by the possibility of randomization. Non-epistemic values need not have been among a scientist’s motivations for choosing one option rather than another if she decided between them by flipping a coin. (This is so even if the decision to flip the coin was itself motivated by non-epistemic considerations.) I read Betz (2013) as proposing this (motivational) version of the randomizer reply in his quick dismissal of inevitability arguments.

Under a third, causal interpretation, inevitability arguments claim that the possession of non-epistemic values necessarily influences choices that are underdetermined by evidence or epistemic

considerations. Causally understood, the randomizer reply denies that the possession of non-epistemic values needs to influence such choices because they can be made randomly. The rejoinder, causally interpreted, points out that even then, the causal impact of non-epistemic values is still felt because the possession of non-epistemic values is what led to the use of randomization in the first place. The choice is still impacted by non-epistemic values, albeit indirectly. The success of this rejoinder to the randomizer reply depends on whether it is true that the decision to randomize is *necessarily* shaped by the possession of non-epistemic values.

Fourth and finally, under a rational interpretation, inevitability arguments claim that non-epistemic values must be involved in the justification of unforced scientific choices. Interpreted in rational terms, the randomizer reply responds that values need not be invoked to defend such choices because they can always be made randomly. (One could read [de Melo-Martín and Intemann \(2016\)](#) as offering something like this version of the randomizer reply.) In my view, this reply succeeds, but only by giving up on justification altogether. A decision made via randomization is one that we have declined to justify ([Neurath 1913](#); [Putnam 2002](#)). It is true that randomization obviates appeal to non-epistemic values in defending a choice, but only because to randomize is to give up such a defense.

The rejoinder to the randomizer reply can also be given a rational interpretation. In the passage quoted above, [Magnus \(2018\)](#) seems to concede that there is some sense in which randomization renders a scientific choice not value-laden. But he can be read as arguing that the decision to randomize itself always requires non-epistemic justification. The reasons that speak in favor of flipping a coin to make a decision, he seems to suggest, are always non-epistemic. Whether the rejoinder succeeds under this interpretation depends on the plausibility of the claim that randomization requires non-epistemic justification, or whether there can instead be exclusively epistemic justifying reasons for randomizing a scientific choice.⁵

Let's grant for the sake of argument that one cannot justify the use of randomization without non-epistemic values, so the decision to randomize a scientific choice is always value-laden in a rational sense. Does it follow that any choice made *via* randomization is also value-laden? Do first-order scientific choices made randomly inherit value-ladenness from the decision to randomize itself? Here we face difficult questions about the relationship between the justification of a choice and the justification of a procedure for making the choice. If authors like [Magnus \(2018\)](#) claim that any scientific choice made randomly must be non-epistemically justified, more of an argument is needed to establish that a decision is rationally value-laden whenever the decision procedure by which it is made is rationally value-laden.

This section has declined to offer a general assessment of inevitability arguments or the randomizer reply. This is in part because there turn out to be multiple versions of each, resting on different senses of value-ladenness. It is also because pulling apart these different versions has surfaced claims that require further argumentation: about whether the decision to randomize is necessarily justified or causally influenced by non-epistemic values, and about how the justification of a choice relates to the justification of the procedure used to make it. Whatever one's position on these issues, I contend that this case study demonstrates the value of the taxonomy characterized above. Developing a sensitivity to different notions of value-ladenness permits a more sophisticated evaluation of claims about value-laden science.

Notes

- 1 Following widespread usage, when discussing values in science, I take a "value-laden" choice to be one that involves *non-epistemic* values.
- 2 Elsewhere Elliott and Korf seem more amenable to the idea that the ontology and roles of values are closely related: "Perhaps attention to these differences [between value concepts] could help to clarify different functional roles that values can play. For example, one could distinguish values that play the functional role of serving as criteria for theory choice from values that serve the functional role of serving as outcomes that would be desirable to achieve" (2024, 22).

- 3 One can think of inevitability arguments as a species of what others have called “gap arguments” (Intemann 2005; Elliott 2011; Brown 2013). However, while these authors often contrast gap arguments with “error arguments,” my characterization of inevitability arguments includes versions of both.
- 4 The randomizer reply is sometimes associated with Neurath’s (1913) discussion of what one should do when one has to make a choice, but there is no way of deciding among the available options on the merits. In such situations, many people adopt “pseudorationalism,” convincing themselves that one of the options is indeed better than the others (9). Neurath recommends instead embracing “the auxiliary motive,” a motive for choosing that does not have to do with the merits of the options. The “purest form” of the auxiliary motive is the drawing of lots (4). Despite this appeal to randomization in the face of underdetermination, it’s a mistake to think of Neurath as advocating the randomizer reply. First, he is considering situations in which one has *no reason* to prefer one option to another, not situations in which one has no *epistemic* reason. Moreover, far from adopting randomization as a way of ridding science of values, Neurath was sympathetic to the idea that values play a role in science (Howard 2006). He considered social and political values to be auxiliary motives and, thus, saw appeal to such values as a legitimate way of making choices in the face of underdetermination.
- 5 Note that I have entertained two different interpretations of proponents’ rejoinder to the randomizer reply: a causal interpretation on which the adoption of randomization is always *causally influenced* by values and a rational interpretation on which randomization must always be *justified* by appeal to non-epistemic values. I don’t think it is clear which of these interpretations Magnus (2018), Douglas (2021), and Havstad (2022) intend.

References

- Baier, Kurt. 1958. *The Moral Point of View: A Rational Basis of Ethics*. Cornell University Press.
- Betz, Gregor. 2013. “In Defence of the Value Free Ideal.” *European Journal for Philosophy of Science* 3 (2): 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Biddle, Justin B. 2013. “State of the Field: Transient Underdetermination and Values in Science.” *Studies in History and Philosophy of Science* 44 (1): 124–33. <http://dx.doi.org/10.1016/j.shpsa.2012.09.003>.
- Biddle, Justin B., and Rebecca Kukla. 2017. “The Geography of Epistemic Risk.” In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Bond, E. J. 1974. “Reasons, Wants and Values.” *Canadian Journal of Philosophy* 3 (3): 333–47. <https://doi.org/10.1080/00455091.1974.10716887>.
- Bond, E. J. 1983. *Reason and Value*. Cambridge University Press.
- Brandon, Robert B. 1998. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. 2nd ed. Harvard University Press.
- Brown, Matthew J. 2013. “Values in Science beyond Underdetermination and Inductive Risk.” *Philosophy of Science* 80 (5): 829–39. <https://doi.org/10.1086/673720>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Davidson, Donald. 1963. “Actions, Reasons, and Causes.” *Journal of Philosophy* 60 (23): 685–700. <https://doi.org/10.2307/2023177>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2016. “The Risk of Using Inductive Risk to Challenge the Value-Free Ideal.” *Philosophy of Science* 83 (4): 500–20. <https://doi.org/10.1086/687259>.
- Dethier, Corey, and Samuel C. Fletcher. 2024. “Consistent Estimators and the Argument from Inductive Risk.” *Paper Presented at the Conference on “Revitalizing Science and Values,”* Center for Philosophy of Science, University of Pittsburgh. April 7, 2024.
- Douglas, Heather E. 2021. *The Rightful Place of Science: Science, Values, and Democracy: The 2016 Descartes Lectures*. Edited by Ted Richards. Consortium for Science, Policy & Outcomes.
- Elabbar, Ahmad. 2023. “The Curatorial View of Assessment and the Ethics of Scientific Advice: Beyond Decisional Autonomy towards Distributive Epistemic Justice.” *International Union for the History and Philosophy of Science and Technology (IUHPST)(Essay Prize in the History and Philosophy of Science)*. https://org/media/pdf/Elabbar_IUHPST_Prize_2023.pdf.
- Elliott, Kevin C. 2011. *Is a Little Pollution Good for You?: Incorporating Societal Values in Environmental Research*. Oxford University Press.
- Elliott, Kevin C., and Rebecca Korf. 2024. “Values in Science: What Are Values, Anyway?” *European Journal for Philosophy of Science* 14 (4): 53. <https://doi.org/10.1007/s13194-024-00615-3>.
- Havstad, Joyce C. 2022. “Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk.” *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.

- Hillgardt, Hannah. 2022. "Looking Beyond Values: The Legitimacy of Social Perspectives, Opinions and Interests in Science." *European Journal for Philosophy of Science* 12 (4): 1–20. <https://doi.org/10.1007/s13194-022-00490-w>.
- Howard, Don. 2006. "Lost Wanderers in the Forest of Knowledge: Some Thoughts on the Discovery-Justification Distinction." In *Revisiting Discovery and Justification: Historical and Philosophical Perspectives on the Context Distinction*, edited by Jutta Schickore and Friedrich Steinle. Springer.
- Hoyningen-Huene, Paul. 1992. "The Interrelations Between the Philosophy, History and Sociology of Science in Thomas Kuhn's Theory of Scientific Development." *British Journal for the Philosophy of Science* 43 (4): 487–501. <https://doi.org/10.1093/bjps/43.4.487>.
- Intemann, Kristen. 2005. "Feminism, Underdetermination, and Values in Science." *Philosophy of Science* 72 (5): 1001–12. <https://doi.org/10.1086/508956>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kuhn, Thomas S. 1977. "Objectivity, Value Judgment, and Theory Choice." In *The Essential Tension: Selected Studies in Scientific Tradition and Change*, Revised ed. University of Chicago Press.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. *Science as Social Knowledge*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Magnus, P. D. 2018. "Science, Values, and the Priority of Evidence." *Logos and Episteme* 9 (4): 413–31. <https://doi.org/10.5840/logos-episteme20189433>.
- McDowell, John. 1996. *Mind and World*. Harvard University Press.
- McMullin, Ernan. 1982. "Values in Science." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982 (2): 2–28. <https://doi.org/10.1086/psaprocbienmeetp.1982.2.192409>.
- Neurath, Otto. 1913. "The Lost Wanderers of Descartes and the Auxiliary Motive." In *Philosophical Papers 1913–1946: With a Bibliography of Neurath in English*, edited by Otto Neurath, Robert Sonn   Cohen, and Marie Neurath. Vienna Circle Collection. Springer.
- Pollard, Bill. 2005. "Naturalizing the Space of Reasons." *International Journal of Philosophical Studies* 13 (1): 69–82. <https://doi.org/10.1080/0967255042000324344>.
- Putnam, Hilary. 2002. "On the Rationality of Preferences." In *The Collapse of the Fact/Value Dichotomy and Other Essays*. Harvard University Press.
- Rooney, Phyllis. 1992. "On Values in Science: Is the Epistemic/Non-Epistemic Distinction Useful?" *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992: 13–22.
- Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton University Press.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Sellars, Wilfred. 1956. "Empiricism and the Philosophy of Mind." In *Minnesota Studies in the Philosophy of Science*, edited by Herbert Feigl and Michael Scriven. University of Minnesota Press.
- Ward, Zina B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science Part A* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Ward, Zina B. 2024. "Disagreement and Values in Science." In *Routledge Handbook on Philosophy of Disagreement*, edited by Maria Baghramian, Adam Carter, and Rach Cosker-Rowland. Routledge.
- Winsberg, Eric. 2012. "Values and Uncertainties in the Predictions of Global Climate Models." *Kennedy Institute of Ethics Journal* 22 (2): 111–37. <https://dx.doi.org/10.1353/ken.2012.0008>.

8

AESTHETIC VALUES IN SCIENTIFIC PRACTICE

Contemporary Debates

Milena Ivanova

8.1 Introduction

In recent years, practitioners from different fields have begun to shed light on the commonalities between art and science and the role played by beauty within their fields. The aesthetics of science has become a fast-growing subfield in philosophy of science, generating important collaborations between philosophers, sociologists, historians, artists, scientists and science communicators. Important contributions in this literature connect art and science by focusing on the notion of representation (Ambrosio and Sánchez-Dorado 2024; Bueno et al. 2018; Frigg and Hunter 2010). Ivanova and French's (2020) collection also explores how beauty affects the evaluation of scientific theories, the relationship between beauty and the acquisition of understanding through models, images and thought experiments and aesthetic experiences going beyond beauty into explorations of the sublime in science. Ivanova and Murphy (2023) further develop novel perspectives on the debate by investigating previously understudied questions in philosophy of science, such as how aesthetic values shape the design, performance and results of experiments, whether there is stability in how the experiment is appreciated aesthetically, what commonalities can be found between artistic and experimental practices and what function negative aesthetic values may play in modern highly collaborative and artificial intelligence (AI) driven experiments.

Several books and articles have also explored important connections between aesthetics and science. McAllister (1996) asks whether practitioners in different scientific traditions agree upon a set of aesthetic values and whether variability of such values may undermine their rational use in science. One of the main ongoing questions in the field is whether beauty can play a positive role in the achievement of epistemic goals such as truth or understanding. While some scholars have argued that aesthetic values can be relevant epistemic factors, as conditions of understanding (Breitenbach 2013; Elgin 2020; Ivanova 2021, 2022; Turner 2019), others have expressed concerns that beauty in science can operate as a cognitive bias stalling scientific progress (Hossenfelder 2018). Examination of aesthetic dimensions of scientific practices has also emerged, giving rise to what I will call “practice-based aesthetics of science.” Wylie (2021) and Turner (2019) pay close attention to the preparation of fossils to understand how aesthetic values shape the preparatory processes and the reception of the final product. The experiment has also received attention in this literature, with new questions emerging. Whether experiments can be aesthetically appreciated has recently been addressed both from a philosophical perspective by Ivanova (2021, 2022, 2023a, 2023b, 2023c; Parsons and Rueger 2000)

as well as a historical perspective (Ball 2005, 2023; Wragge-Morley 2020). In sociology, studies on wellbeing in science are showing an interesting link between scientists' aesthetic experiences in daily practice and their overall sense of purpose, accomplishment and wellbeing (Jacobi et al. 2023; Jacobi, Varga, and Vaidyanathan 2022; Vaidyanathan, Haraburda, and Jacobi 2023). These works shed new light on the classic questions regarding the goal of science and the role of aesthetic values in it but have also opened a plethora of new directions for exploration, particularly when it comes to different scientific practices and experiences scientists encounter in their work.

In this chapter, I reflect on the recent developments in the aesthetics of science literature, identify emerging directions and draw important analogies with the broader literature on values in science. In Section 8.2, I show the breadth of contexts in which we can see aesthetic values operating in science and their role in scientific reasoning. In Section 8.3, I present an overview of arguments defending the cognitive role of aesthetic values, and the counterargument takes these values to operate as cognitive biases. Section 8.4 presents what I call the “practice turn” in the aesthetics of science, where I show novel developments in the literature asking questions about the role of aesthetics in the daily work of scientists. In Section 8.5, I offer my reflections on the new directions emerging in this field and the future of beauty in science. Section 8.6 is the conclusion.

8.2 Aesthetic values in science: Where do we find them and what do they do?

We can distinguish between three different levels in which aesthetic judgements are made in science: the objects of scientific inquiry, the products created by scientists and the processes by which these products are made (Ivanova 2017a, 2020). Starting with natural phenomena, entities and processes, the objects of scientific enquiry can exhibit aesthetic properties and evoke aesthetic experiences. From rosy bonnet mushrooms, rainbow lorikeets and hexagonal honeycombs, to sunsets and waterfalls, nature is a source of aesthetic experiences (Brady and Prior 2020; Parsons 2023). Aesthetic experiences are also afforded when engaging with nature under carefully crafted experimental settings, such as the observation of crystallisation under a powerful microscope or the decomposition of light through a prism.

The products scientists create in their work are also subject to aesthetic appreciation. Scientific theories, experiments, images, models and instruments (Ivanova 2017a, 2024; Ivanova and French 2020; Ivanova and Murphy 2023; McAllister 1996) as well as mathematical proofs (Dutilh Novaes 2019; Kuipers 2002; Montano 2014) are often praised for their beauty, elegance, symmetry, simplicity and unity. For instance, Newtonian mechanics can be seen as a highly beautiful theory due to its elegance in describing the motion of bodies with three rather simple laws of motion and a law of gravitation featuring few parameters. Einstein's theories of special and general relativity are regarded as beautiful due to their unifying explanatory power achieved through a small set of scientific principles (Chandrasekher 1987; Ivanova 2020). Galileo Galilei's thought experiment on falling bodies is among the most beautiful thought experiments. As Murphy (2020) argues, this experiment utilises economic concepts and materials to demonstrate an important novel concept: that of acceleration. Scientific images also have aesthetic value. Take the images NASA publicly disseminated in 2021, produced with the James Webb Space Telescope, which were the focus of much discussion due to their beauty. To produce the images, scientists used aesthetic values when translating the data collected by the telescope, for instance, selecting colours not detected by the telescope to present the data (Ivanova 2022).

In addition to the objects of study and products of scientific activity, the very process of developing a theory, deriving a mathematical proof, designing and performing a scientific experiment or preparing a scientific image, involves the creativity, imagination and aesthetic sensibility of scientists. We can draw parallels between these scientific practices and artistic productions, seeing them

as involving creative agency, aesthetic values and other goals of the scientist (Anscomb 2023). In fact, such a connection can be found in scientific work today just as much as in earlier traditions. Going back to the production of NASA's images, where astronomers made choices in the selection of colours when translating the raw data collected by the telescope into images accessible to the human eye, we can similarly see aesthetic values in the preparation of images in other scientific traditions. For instance, Leonardo da Vinci in anatomy and Robert Hooke in entomology would enhance the aesthetic features of their subject matter in their depictions, generating in the viewer awe and wonder when engaging with the depicted specimen. Today scientists often use aesthetic values when performing and designing experiments and images. Aesthetic judgements are integral to many scientific fields, from astronomy to biology and medicine to chemistry and palaeontology. The very process of scientific discovery is also often compared to artistic production. The French physicist, historian and philosopher of science, Pierre Duhem, argues that it is impossible to follow the process of constructing a physical theory "without feeling keenly that such a creation of the human mind is truly a work of art" (1906, 24). Ernest Rutherford similarly claims that "the process of scientific discovery may be regarded as a form of art," continuing that "a well constructed theory is in some respects undoubtedly an artistic production" (quoted in McAllister 1996, 14). In addition to seeing the creative process in science as parallel to artistic creativity, Henri Poincaré explicitly devises an account of scientific creativity at the heart of which is the aesthetic sensibility of the scientist, which acts as "the delicate sieve" that scans and evaluates which ideas that the mind produces are useful by judging first their aesthetic appeal. At the heart of this account is the idea that "care for the beautiful is care for the useful," that the elegant, coherent and beautiful hypotheses, experiments or proofs are likely to prove more fruitful to the scientists in their inquiry than the less coherent, disjointed or overcomplex ones (Ivanova 2017b).

So far I have illustrated the rich aesthetic dimension of science, but what is the function of these judgements? While aesthetic values are often seen as a motivator in scientific inquiry, some believe aesthetic values are epistemically significant, in facilitating understanding or indicating truth. To illustrate the motivational role of aesthetic values, the French mathematician Henri Poincaré argues that "[t]he scientist does not study nature because it is useful to do so. He studies it because he takes pleasure in it, and he takes pleasure in it because it is beautiful" (2001, 368). Poincaré's view on the role of aesthetic values in science, however, goes beyond seeing these values purely as motivational. Rather, aesthetic values are constitutive of our achievement of understanding:

I am not speaking, of course, of the beauty which strikes the senses, of the beauty of qualities and appearances. I am far from despising this, but it has nothing to do with science. What I mean is that more intimate beauty which comes from the harmonious order of its parts, and which pure intelligence can grasp.

(2001, 368)

The Nobel laureate Subrahmanyan Chandrasekhar's book *Beauty and Truth* further offers a series of arguments to convince that beauty guides physicists to the truth, claiming that "in the arts as in the sciences, the quest is for the very same elusive quality: beauty" (1987, 52). During the early days of Einstein's general theory of relativity, many expressed the belief that the theory's beauty offered support for its truth. For instance, Paul Dirac claimed that "one has a great confidence in [a] theory arising from its great beauty, quite independently of its detailed successes" (1980, 40). Before Arthur Eddington's expeditions in 1919 that offered empirical support for the theory, Dirac argued that "[o]ne has an overpowering belief that [the theory's] foundations must be correct quite independently of its agreement with observation" (1980, 40). This conviction was shared by many physicists at the time, including Eddington himself as well as Werner Heisenberg, who claimed that "[i]f nature leads

us to mathematical forms of great simplicity and beauty we cannot help thinking that they are ‘true,’ that they reveal a genuine feature of nature” (1971, 68). In the following section, I focus on arguments defending, and arguing against, the epistemic significance of aesthetic values.

8.3 Beauty: A guide to the truth or systemic bias?

One of the important issues in the literature on science and values has been the question of how, if at all, to distinguish epistemic or cognitive values from non-epistemic or contextual values (see e.g., Douglas 2013; Longino 1996; Rooney 2017; Steel 2010). In the context of aesthetic values, some have explored whether aesthetic values are epistemic. How can we justify the idea that beauty leads to the truth? One way is to argue that there is evidence from past practices that being guided by aesthetic value leads to empirical success or approximate truth. Chandrasekhar argues that “we have evidence that a theory developed by a scientist, with an exceptionally well-developed aesthetic sensibility, can turn out to be true even if, at the time of its formulation, it appeared not to be so” (1987, 64). Chandrasekhar argues that uncovering the beauty of nature is a primary motivation in science and a guide to the truth. Focusing on the beauty of Einstein’s theory of general relativity, to be found in the unification of the fundamental concepts of space and time, matter and motion, with elegance and beauty, he claims that such values are to be trusted because they have led us to the truth many times before.

Another track record argument can be found in James McAllister’s *Beauty and Revolution in Science* (1996). Rather than defending the idea that beauty leads to the truth, McAllister focuses on the empirical adequacy of scientific theories and argues that aesthetic values can be good predictors of such success. McAllister argues that scientists form aesthetic canons, a set of aesthetic values reflected in successful theories adopted at the time and their predecessors. When a new theory comes along, even if the theory has yet to receive confirmation, scientists can rationally make an aesthetic induction from the past and infer whether a new theory is to be trusted on the basis of its aesthetic features, given that those are constitutive of the adopted aesthetic canon. For McAllister, however, aesthetic cultures and canons are unstable. The history of science is a history of the evolution of the concept of beauty, with new empirically successful theories revising and transforming the aesthetic canon that communities adopt.

Ulianov Montano (2014) challenges the prescriptive accuracy of this framework, pointing to cases where epistemically successful theories or mathematical proofs are adopted by communities but are met with resistance because they do not satisfy the scientists’ aesthetic sensibilities. McAllister’s theory predicts that with time, scientists learn to appreciate the traits of successful theories and revise their aesthetic canon. Montano argues that this is not the case, pointing to the case of computer-assisted proofs in mathematics, where, despite their success, such proofs continue to be considered ugly. Ivanova (2020) further points to the case of quantum mechanics and the standard model, which, despite being highly successful, have not resulted in the community revising their aesthetic canons and continue to be regarded as displeasing.

Another issue with track-record-based justifications is cases of aesthetically pleasing theories that have turned out to be false. Projects in high-level physics, such as the Kaluza-Klein theory, provide a beautiful way of unifying gravity and the other gauge fields, but are false. In *Lost in Math: How Beauty Leads Physics Astray* (2018), Sabine Hossenfelder reflects on such cases, arguing that beauty can operate as a systematic bias in a community, leading to unproductive projects and stagnation of ideas. On the other hand, as noted earlier, we also see that successful theories can violate our aesthetic expectations—as is the case of the standard model and quantum mechanics. Ivanova (2020, 2024) argues that such inductive arguments are ultimately not going to convince when it comes to the epistemic role of aesthetic values, since inductive arguments pulling in both directions can be made.

Thus, the recent move of identifying the cognitive role of aesthetic values not as truth tracking but as conditions for understanding took place.

Motivated by the move in philosophy of science to study how models represent the world via omission and additions of features not present in the target system, Catherine Elgin (2020), among others, argues for a non-factual account of understanding. The goal of science is not to deliver true propositions but to allow us to understand how to manipulate and interact with systems we are interested in. The move towards non-factual understanding presents a shift in the debate on the goals of science, and proponents have offered their reflections on the role of aesthetic factors in achieving this epistemic goal. Elgin (2020) argues that aesthetic values are a constitutive part of our desire to gain understanding of the world and operate as “gatekeepers,” leading us to productive ideas. This is not to say that these are non-revisable and fixed no matter what; these values can come into conflict with other values we hold, but their role in the acquisition of understanding is central. Ivanova (2017a, 2020) further argues that aesthetic values are conditions of human understanding, reflecting Poincaré’s insightful account according to which “[i]n formulating a general, simple, and formal law, based on a comparatively small number of not altogether consistent experiments, we have only obeyed a necessity from which the human mind cannot free itself” (Poincaré 2001, 100). This aesthetic experience, the “harmony” we experience when we reflect on our theories “is at once a satisfaction of our aesthetic requirements, and an assistance to the mind which it supports and guides” (Poincaré 2001, 396–7). Similar insights also come from Kantian scholarship, where aesthetic values are seen as conditions of understanding (Breitenbach 2013). Empirical studies also seem to support the idea that aesthetic values constitute human reasoning and understanding, with Tania Lombrozo (2016) arguing that

when children and adults generate and evaluate explanations, they recruit explanatory virtues, such as simplicity and breadth, as evaluative constraints on reasoning. As a result, they are more likely to generate and favor broad and simple hypotheses, and to discover broad and simple patterns. (749)

Our desire for simplicity, unity and elegance then affects both our preferences for what hypotheses we construct and what patterns in nature we focus on. This is not to say that if systematically in error, people resist revising their ideas, but that there is a barrier to accepting complex and messy explanations. So far, we have reflected on the justification of aesthetic values when it comes to epistemic goals, such as truth, empirical adequacy and understanding. However, in the very recent literature, we are seeing attention being paid to the daily practices of scientists and a desire to understand how aesthetic values shape these activities. The next section explores this recent shift.

8.4 The turn to scientific practice

In *The Neglect of Experiment* (1989), Allan Franklin points to the important epistemological questions arising in experimental practice that had been left out of philosophy of science due to the narrow focus on theory over practice, leaving the autonomous life of experiments understudied. I voice similar concerns with the aesthetics of science literature, where much attention has been paid to the role of aesthetic values in the evaluation of scientific theories, leaving much of scientific practice understudied. Yet recently, we are seeing a move towards what I will call “practice-focused aesthetics of science.” Recent works focus on the aesthetic judgement at play in the preparation of fossils and the speculative reasoning they afford (Currie 2023; Turner 2019; Wylie 2021), chemical biology, including the beauty of molecules, molecular images and synthesis (Parsons 2012), anatomy (Clarke and Ambrosio 2018) and thought experiments (Murphy 2020). Crucially, we have seen the emergence of

a debate dedicated to the aesthetics of experimental practice (Ball 2005, 2023; Parsons and Rueger 2000; Ivanova 2021, 2023a, 2023b, 2023c, 2024; Ivanova and Murphy 2023; Ivanova et al. 2023; Murphy, Currie and Walsh forthcoming).

In “What is a Beautiful Experiment” (2023b), Ivanova offers a new framework to understand the aesthetic value of experiments. Several aesthetic aspects of the experiment are identified: the subject matter, the instruments employed, the design of the experiment, its performance, its results and the process of creating it. While all these aspects can involve the aesthetic sensibility, Ivanova draws a distinction between *perceptual beauty*, the visual or tactile beauty of an eighteenth-century microscope, for instance, and *intellectual beauty*, experienced when we appreciate how the design of an experiment connects to its significance. A beautiful experiment, according to Ivanova (2021, 2023b), is one that has an optimal design and produces significant results, where the latter are understood in a pluralistic way—they can align with expectation by confirming or discovering something theoretically expected, or by producing disruptive results that spark new investigations.

Attention is also being paid to the different kinds of aesthetic experiences associated with science. Arcangeli and Dokic (2020), for instance, offer an illuminating categorisation of aesthetic experiences in science, arguing that what we consider “positive” aesthetic experiences, usually associated with pleasure, beauty, proportion, elegance and symmetry, are due to fluency processing, ease of understanding the subject matter. Disfluent processing, on the other hand, is associated with more sublime experiences; those by their nature have both a positive and a negative component. Sublime experiences point to our limitations, smallness and can be experienced as disruptive and negative. They are, however, cognitively significant, since they can illuminate the limitations of our theoretical frameworks and thus help us expand them. Murphy (2023) further focuses on the value of profundity. Focusing on thought experiments such as Schrodinger’s cat, Murphy argues that thought experiments can be aesthetically valuable because of their ability to convey important ideas with minimal materials. Murphy demonstrates the scope to broaden the class of aesthetic values that are studied in science and highlights a potential limitation in the fluency processing account of the aesthetic experience in science. While fluency can help us explain the presumed connections between the aesthetic and the epistemic, profundity can align more with the disfluency account, opening space for a more pluralistic account of aesthetic experiences in science and their cognitive dimension.

Moving towards negative aesthetic values, Sophie Ritson (2023) studies ugly experimental results. In high-energy physics, where physicists are working on theories beyond the standard model, Ritson argues that scientists are explicitly looking for ugly, non-conforming results because they believe such surprising and unexplainable results can push them in directions. Similarly, Stuart (2023) explores experiments designed using AI and argues that these hold great epistemic value by not conforming to our aesthetic values. Seeing aesthetic values as limiting the set of possibilities human scientists can consider, Stuart argues that AI-designed ugly experiments can generate epistemic goods. In the latter works, we see the epistemic and aesthetic connection receiving new analysis, where negative aesthetic values are seen as epistemically significant too.

8.5 New directions in the aesthetics of science

One promising direction emerging in the literature concerns the connections between art and science and aesthetics and philosophy of science. While these analogies took central stage a couple of decades ago when it came to the development of the literature on scientific models and representation, currently, philosophers of science are turning to broader explorations and analogies. For one, attention is paid to the diverse aesthetic experiences and values in science, from sublime (Arcangeli and Dokic 2020) and profundity (Murphy 2023), to negative aesthetic values such as ugliness (Ritson 2023; Stuart 2023). In addition to this, the idea of aesthetic cultures is receiving new attention.

McAllister's work explores aesthetic cultures and revolutions by focusing on theorising. Parsons and Reuger (2000) argue that aesthetic appreciation of experiments changes when methodological changes in science occur, giving the experiment a different role and thus placing a different focus on its aesthetic value. Ivanova (2023) on the contrary argues that we can see stability in how experiments are aesthetically appreciated, specifically when it comes to appreciating the beauty in a well-designed experiment with significant results. Anscomb (2023) appeals to "functional beauty" to understand the beauty of experimental art in the twentieth century, much of which focuses on the rejection of beauty. Despite the dadaist movement divorcing itself from the idea of visually pleasing beauty, Anscomb argues that such works can still be functionally beautiful, which we appreciate by reflecting on how an artist has achieved their goal with the materials of choice, rather than the visual features of the artwork. Anscomb makes the case that the concept of functional beauty accommodates the aesthetic value not only of experimental art but also of scientific experiments. Ivanova (2023) further argues that such a conception can also be used in contemporary large-scale experiments where there is a sense of distance between the performance of the experiment and the observation of a result, appreciating the relationship between the design and the result can unveil its "intellectual beauty."

Another important area of emerging research concerns the use of AI in creative labour and the role of creative agency in scientific work. Stuart (2023) discusses a case from quantum mechanics in which AI assists in finding novel solutions to experimental configurations. As discussed above, being unconstrained by our aesthetic requirements, argues Stuart, AI can offer scientists a broader range of possibilities that will nevertheless be aesthetically displeasing. Ivanova (2024) argues that in specific contexts, AI could enhance the creative possibilities of scientists, but there are also concerns that AI ultimately produces convergent ideas, which in the long term could compromise the pursuit of radically novel ideas, often seen as constitutive of scientific progress (Wadinambiarachchi et al. 2024). Ivanova et al. (2023) also reflect on the sense of creative agency for broader scientific goals, such as flourishing and wellbeing. Their study shows that scientists who feel creatively involved in their work feel a higher sense of aesthetic value and wellbeing. Murphy et al. (forthcoming) also offer an illuminating case for considering some experiments in science aesthetically valuable due to their affording aesthetic experiences in the practitioner to do with their own sense of agency and embodiment in the performance of the experiments, calling for further attention to be paid to this aesthetic aspect of experimental work. These aspects posit new and understudied implications for the involvement of AI in scientific work and open further questions for exploration on the relationship between scientific progress, aesthetic values and wellbeing in science.

Beyond the new topics emerging in the literature, we also see a shift towards the use of diverse methodologies and interdisciplinary engagements. The core work in this topic had previously primarily engaged with historical case studies or scientists' own reflections on their work, while the more recent literature shows diversification of methods, with quantitative and qualitative methods being used. These methods illuminate contemporary thinking among scientists today and complement theoretical frameworks arrived at by case studies. One example of such work is Stuart's ethnographic work, looking at the role of creative agency in one lab, closely interviewing four practitioners and studying how their positionality, career stage and gender affect their perception of the role of creativity in their work. Another example is the work of Vaidyanathan, Haraburda, and Jacobi (2023), offering large quantitative data on the aesthetic experiences of over 3000 biologists and physicists in four countries and qualitative data from detailed interviews with 215 scientists. Data from this study have supported the theoretical framework on the aesthetic levels of scientific experiments developed by Ivanova (2023), which had been arrived at by examining historical case studies. In their collaborative work, Ivanova et al. (2024) offer further support for this framework, showing that contemporary scientists identify the aesthetic significance of experiments in the same ways. The quantitative study also shows some interesting places of deviation, opening further questions for

exploration, such as the role of agency in contemporary experimental settings. Furthermore, more recently, [Ivanova and Vaidyanathan \(2024\)](#) offer a further study on the aesthetic nature of surprise in science, also based on qualitative data. Many philosophers of science have been defending the use of empirical methods as complementary to traditional case studies, including [Nersessian \(2009\)](#), [Stuart \(2019, 2023\)](#), [Kozlov \(2023a, 2023b\)](#), [Schindler \(2022\)](#), [Wilkenfeld and Lombrozo \(2020\)](#), and [Hangel and ChoGlueck \(2023\)](#). Such methodologically diverse studies are emerging in the aesthetics of science and promise to offer support for existing theoretical frameworks and identify further areas for research.

8.6 Conclusion

The debate on values in science has traditionally been divided by those who advocate for the “value free ideal,” raising awareness that individual values can operate as cognitive biases that harm science, and those who argue that science cannot and should not be value-free. Feminist philosophers of science, for one, have pointed to the values a community ought to employ to identify and amend harmful individual biases, such as equality among practitioners and facilitation of democratic discourse ([Longino 1990](#)). The debate has been primarily framed around the moral and political values in science, and in this debate, aesthetic values are missing.

What can we learn from the above discussion about the place and role of aesthetic values of science for this broader discourse? Some worry that beauty might operate as a bias just like gender bias, stifling scientific progress. I think we will be missing something important about scientific practice; however, if we understand aesthetic values in science as harmful biases. What we learn from the aesthetics of science is that there is more to these values and experiences that science affords us. Aesthetic values are held closely by scientists; they can lead to productive ideas and yet sometimes also mislead. We can see from ongoing debates in particle physics that the community is reflecting critically on their aesthetic preferences and considering whether to pursue aesthetically disvalued ideas in the goal of broadening possibilities, which aligns with how the values literature more broadly has been debating the conditions under which particular value influences are productive or unproductive in science ([Holman and Wilholt 2022](#); [Resnik and Elliott 2023](#)).

But the aesthetics of science literature also studies why scientists persist in the pursuit of certain hypotheses, scientists’ daily practices in the lab or in the field and the rich aesthetic life that acts as a motivator and guide in the pursuit of understanding the world. This could be a model for the broader field of science and values to consider the wide variety of ways that values relate to scientific practice. Building on the argument from inductive risk ([Douglas 2009](#); [Elliott and Richards 2017](#)), there has been a tendency for this literature to focus primarily on how values influence hypothesis assessment, but contemporary discussions of aesthetic values illustrate the benefits of considering a wider variety of ways that values can relate to scientific practice. This chapter demonstrates how rich scientists’ aesthetic life is by identifying the levels at which aesthetic values are experienced and their connection to other scientific goals that make scientific activities possible.

References

- Ambrosio, Chiara, and Sánchez-Dorado, Julia (eds.) (2024). *Abstraction in Science and Art: Philosophical Perspectives*. Routledge.
- Anscomb, Claire. 2023. “Beautiful Experiments in Art and Science.” In *The Aesthetics of Scientific Experiments*, edited by Milena Ivanova and Alice Murphy. Routledge.
- Arcangeli, Margherita, and Jérôme Dokic. 2020. “A Plea for the Sublime in Science.” In *The Aesthetics of Science: Beauty, Imagination, Understanding*, edited by Milena Ivanova and Steven French. Routledge.
- Ball, Philip. 2005. *Elegant Solutions: Ten Beautiful Experiments in Chemistry*. The Royal Society of Chemistry.

- Ball, Philip. 2023. *Beautiful Experiments: An Illustrated History of Experimental Science*. University of Chicago Press.
- Brady, Emily, and Jonathan Prior. 2020. "Environmental Aesthetics: A Synthetic Review." *People and Nature* 2: 254–66. <https://doi.org/10.1002/pan3.10089>.
- Breitenbach, Angela. 2013. "Aesthetics in Science: A Kantian Proposal." *Proceedings of the Aristotelian Society* 113 (1): 83–100. <https://doi.org/10.1111/j.1467-9264.2013.00346.x>.
- Bueno, Otávio, George Darby, Steven French, and Dean Rickles. 2018. *Thinking About Science, Reflecting on Art*. Routledge.
- Chandrasekher, Subrahmanyan. 1987. *Truth and Beauty: Aesthetics and Motivation in Science*. University of Chicago Press.
- Clarke, Brendan, and Chiara Ambrosio. 2018. "The Nervous System and the Anatomy of Expression: Sir Charles Bell's Anatomical Watercolours." In *Imagining the Brain: Episodes in the History of Brain Research*, edited by Chiara Ambrosio and William MacLehose. Elsevier.
- Currie, Adrian. 2023. "Epistemic Engagement, Aesthetic Value and Scientific Practice." *British Journal for Philosophy of Science* 74 (2). <https://doi.org/10.1086/714802>.
- Dirac, Paul. 1980. "The Excellence of Einstein's Theory of Gravitation." In *Einstein: The First Hundred Years*, edited by Maurice Goldsmith, Alan Mackay, and James Woudhuysen. Pergamon Press.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2013. "The Value of Cognitive Values." *Philosophy of Science* 80 (5): 796–806. <https://doi.org/10.1086/673716>.
- Duhem, Pierre. 1906. *The Aim and Structure of Physical Theory*. Princeton University Press.
- Dutilh Novaes, Catarina. 2019. "The Beauty (?) of Mathematical Proofs." In *Advances in Experimental Philosophy of Logic and Mathematics*, edited by Andrew Aberdein and Matthew Inglis. Bloomsbury.
- Elgin, Catherine Z. 2020. "Epistemic Gatekeepers: The Role of Aesthetic Factors in Science." In *The Aesthetics of Science: Beauty, Imagination and Understanding*, edited by Milena Ivanova and Steven French. Routledge.
- Elliott, Kevin C., and Ted Richards (eds.) 2017. *Exploring Inductive Risk: Case Studies of Values in Science*. Oxford University Press.
- Franklin, Allan. 1989. *The Neglect of Experiment*. Cambridge University Press.
- Frigg, Roman, and Matthew Hunter. 2010. *Beyond Mimesis and Convention: Representation in Art and Science*. Springer.
- Hangel, Nora, and Christopher ChoGlueck. 2023. "On the Pursuitworthiness of Qualitative Methods in Empirical Philosophy of Science." *Studies in History and Philosophy of Science* 98: 29–39. <https://doi.org/10.1016/j.shpsa.2022.12.009>.
- Heisenberg, Werner. 1971. *Physics and Beyond: Encounters and Conversations*, translated by Arnold J. Pomerans. George Allen & Unwin.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Hossenfelder, Sabine. 2018. *Lost in Math: How Beauty Lead Physicists Astray*. Basic Books.
- Ivanova, Milena. 2017a. "Aesthetic Values in Science." *Philosophy Compass* 12 (10): e12433. <https://doi.org/10.1111/phc3.12433>.
- Ivanova, Milena. 2017b. "Poincaré's Aesthetics of Science." *Synthese* 194: 2581–94. <https://doi.org/10.1007/s11229-016-1069-1>.
- Ivanova, Milena. 2020. "Beauty, Truth and Understanding." In *The Aesthetics of Science: Beauty, Imagination and Understanding*, edited by Milena Ivanova and Stephen French. Routledge.
- Ivanova, Milena. 2021. "The Aesthetic of Experiments." *Philosophy Compass* 16 (3): e12730. <https://doi.org/10.1111/phc3.12730>.
- Ivanova, Milena. 2022. "James Webb Images Were Made to Look Beautiful." *IAI News*, August 22. <https://iai.tv/articles/james-webb-images-were-made-to-look-beautiful-auid-2222>.
- Ivanova, Milena. 2023a. "Scientific Progress and Aesthetic Values" In *New Philosophical Perspectives on Scientific Progress*, edited by Yefeng Shan. Routledge.
- Ivanova, Milena. 2023b. "What Is a Beautiful Experiment?" *Erkenntnis* 88: 3419–37. <https://doi.org/10.1007/s10670-021-00509-3>.
- Ivanova, Milena. 2023c. "The Aesthetic Value of Scientific Experiments." In *The Aesthetics of Scientific Experiments*, edited by Milena Ivanova and Alice Murphy. Routledge.
- Ivanova, Milena. 2024. "Aesthetic Virtues and Theory Acceptance." In *Scientific Theories and Philosophical Stances: Themes from van Fraassen*, edited by Claus Beisbart and Michael Frauchiger. De Gruyter. <https://doi.org/10.1515/9783111019802-010>.

- Ivanova, Milena, and Alice Murphy (eds.). 2023. *The Aesthetics of Scientific Experiments*. Routledge.
- Ivanova, Milena, and Brandon Vaidyanathan. 2024. "Surprise in Science: A Qualitative Study." *Erkenntnis*. <https://doi.org/10.1007/s10670-024-00880-x>.
- Ivanova, Milena, and Stephen French (eds.). 2020. *The Aesthetics of Science: Beauty, Imagination, Understanding*. Routledge.
- Ivanova, Milena, Bridget Ritz, Marcela Duque, and Brandon Vaidyanathan. 2024. "Beauty in Experiment: A Qualitative Analysis of Aesthetic Experiences in Scientific Practice." *Studies in History and Philosophy of Science* 104: 3–11. <https://doi.org/10.1016/j.shpsa.2024.02.003>.
- Jacobi, Christopher, Peter J. Varga, and Brandon Vaidyanathan. 2022. "Aesthetic Experiences and Flourishing in Science: A Four-Country Study." *Frontiers in Psychology* 13: 923940. <https://doi.org/10.3389/fpsyg.2022.923940>.
- Jacobi, Christopher, Peter J. Varga, Zohaib Jessani, and Brandon Vaidyanathan. 2023. "Individual Differences in Scientists' Aesthetic Disposition, Aesthetic Experiences, and Aesthetic Sensitivity in Scientific Work." *Frontiers in Psychology* 14: 1197870. <https://doi.org/10.3389/fpsyg.2023.1197870>.
- Kozlov, Anatolii. 2023a. "Emotions in Scientific Practice." *Interdisciplinary Science Reviews* 48 (2): 329–48. <https://doi.org/10.1080/03080188.2023.2193073>.
- Kozlov, Anatolii. 2023b. "Scientific Experiments beyond Surprise and Beauty." *European Journal for Philosophy of Science* 13: 38. <https://doi.org/10.1007/s13194-023-00536-7>.
- Kuipers, Theo A. F. 2002. "Beauty, A Road to the Truth." *Synthese* 131: 291–328. <https://doi.org/10.1023/A:1016188509393>.
- Lombrozo, Tania. 2016. "Explanatory Preferences Shape Learning and Inference." *Trends in Cognitive Sciences* 20 (10): 748–59.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Longino, Helen E. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn Hankinson Nelson and Jack Nelson. Kluwer.
- McAllister, James W. 1996. *Beauty and Revolution in Science*. Cornell University Press.
- Montano, Uliyanov. 2014. *Explaining Beauty in Mathematics: An Aesthetic Theory of Mathematics*. Springer.
- Murphy, Alice. 2020. "The Aesthetic and Literary Qualities of Scientific Thought Experiments." In *The Aesthetics of Science: Beauty, Imagination, Understanding*, edited by Milena Ivanova and Steven French. Routledge.
- Murphy, Alice. 2023. "Profound Experiments." In *The Aesthetics of Scientific Experiments*, edited by Milena Ivanova and Alice Murphy. Routledge.
- Murphy, Alice, Adrian Currie, and Kirsten Walsh. Forthcoming. "Aesthetics and Agency in Experiments." *Australasian Journal of Philosophy*.
- Nersessian, Nancy J. 2009. "How Do Engineering Scientists Think? Model-Based Simulation in Biomedical Engineering Research Laboratories." *Topics in Cognitive Science* 1 (4): 730–57. <https://doi.org/10.1111/j.1756-8765.2009.01032.x>.
- Parsons, Glenn. 2012. "The Aesthetics of Chemical Biology." *Current Opinion in Chemical Biology* 16 (5–6): 576–80. <https://doi.org/10.1016/j.cbpa.2012.10.025>.
- Parsons, Glenn. 2023. *Aesthetics and Nature: The Appreciation of Natural Beauty and the Environment*. Bloomsbury.
- Parsons, Glenn G., and Alexander Rueger. 2000. "The Epistemic Significance of Appreciating Experiments Aesthetically." *British Journal of Aesthetics* 40 (4): 407–23. <https://doi.org/10.1093/bjaesthetics/40.4.407>.
- Poincaré, Henri. 2001. *The Value of Science: Essential Writings of Henri Poincaré*, edited by Stephen Jay Gould. Modern Library.
- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54: 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Ritson, Sophie. 2023. "Something from Nothing: 'Non-discovery' and Transformations in High Energy Experimental Physics at the Large Hadron Collider." In *The Aesthetics of Scientific Experiments*, edited by Milena Ivanova and Alice Murphy. Routledge.
- Rooney, Phyllis. 2017. "The Borderlands between Epistemic and Non-Epistemic Values." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Schindler, Samuel. 2022. "Theoretical Virtues: Do Scientists Think What Philosophers Think They Ought to Think?" *Philosophy of Science* 89 (3): 542–64. <https://doi.org/10.1017/psa.2021.40>.
- Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. <https://doi.org/10.1086/650206>.
- Stuart, Michael T. 2019. "Everyday Scientific Imagination: A Qualitative Study of the Uses, Norms, and Pedagogy of Imagination in Science." *Science & Education* 28: 711–30. <https://doi.org/10.1007/s11191-019-00067-9>.

- Stuart, Michael T. 2023. "The Future Won't Be Pretty: The Nature and Value of Ugly, AI Designed Experiments." In *The Aesthetics of Scientific Experiments*, edited by Milena Ivanova and Alice Murphy. Routledge.
- Turner, Derek. 2019. *Paleoaesthetics and the Practice of Paleontology*. Cambridge University Press.
- Vaidyanathan, Brandon, Bailey Haraburda, and Christopher Justin Jacobi. 2023. "Beauty in Biology: An Empirical Assessment." *Journal of Biosciences* 48: 15. <https://doi.org/10.1007/s12038-023-00342-6>.
- Wadinambiarachchi, Samangi, Ryan M. Kelly, Saumya Pareek, Qiushi Zhou, and Eduardo Velloso. 2024. "The Effects of Generative AI on Design Fixation and Divergent Thinking." *CHI '24: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. Article No. 380: 1–18. <https://doi.org/10.1145/3613904.3642919>.
- Wilkenfeld, Daniel A., and Tania Lombrozo. 2020. "Explanation Classification Depends on Understanding: Extending the Epistemic Side-Effect Effect." *Synthese* 197: 2565–92. <https://doi.org/10.1007/s11229-018-1835-3>.
- Wragge-Morley, Alexander. 2020. *Aesthetic Science: Representing Nature in the Royal Society of London, 1650–1720*. The University of Chicago Press.
- Wylie, Caitlin Donahue. 2021. *Preparing Dinosaurs: The Work behind the Scenes*. MIT Press.

9

REPRESENTATIONAL RISK REVISITED

Stephanie Harvard and Eric Winsberg

9.1 Chapter overview

In this chapter, we introduce the concept of “representational risk.” In the next section, we describe the concept and explain why we think it is useful to distinguish representational risk from inductive risk. In [Section 9.3](#), we analyze representational risk using a simplified health economics model, focusing on the nature of representational decisions and the unique downstream harms that can result from them. [Section 9.4](#) concludes by pointing to future questions philosophers could ask about representational risk.

9.2 From inductive risk to representational risk

Richard [Rudner \(1953\)](#) famously claimed that science inevitably involves values because scientists need to accept or reject hypotheses—and since scientific evidence almost never affords certainty, it is always possible that scientists will mistakenly accept a false hypothesis or reject a true one. According to Rudner, then, deciding when to accept/reject a hypothesis requires scientists to make an ethical judgment: *How serious would it be to make a mistake?* If scientists think it would be a relatively serious (or relatively minor) ethical mistake to accept a false hypothesis, they will demand more evidence (or less evidence) before accepting, and so on. In [Rudner’s \(1953\)](#) terms: “Obviously our decision regarding the evidence and respecting how strong is ‘strong enough’, is going to be a function of the importance, in the typically ethical sense, of making a mistake in accepting or rejecting the hypothesis” (2). This is a persuasive account of why science is not value-free.

Carl [Hempel \(1954\)](#) was the first to use the term “inductive risk” to point to something in the neighborhood of what Rudner was talking about. Specifically, [Hempel \(1954\)](#) remarked that scientists can err in making generalizations. For example, scientists might attribute the disposition “solubility-in-water” to a lump of sugar that is not actually put in water, even though to do so “is to make a generalization, and this involves an inductive risk” (14). In a later work, [Hempel \(1965\)](#) emphasized that any *law* in empirical science must be accepted on the basis of incomplete evidence, and that acceptance, therefore, “carries the ‘inductive risk’ that the presumptive law may not hold in full generality” (92). [Hempel \(1965\)](#) further noted that when “a scientific rule of acceptance” is applied to a hypothesis on the basis of evidence, there are four possible outcomes. Accepting a true hypothesis and rejecting a false hypothesis are the two outcomes that science aims to achieve, while the other two possibilities represent “the inductive risk that any acceptance rule must involve” (92).

In an influential paper, Douglas (2000) connected Hempel's (1965) concept of inductive risk to previous philosophical work on the role of values in science, including Rudner's (1953). Indeed, Douglas (2000) asserted that between 1948 and 1965 "All those that argued for the legitimate use of values in science did so on the basis of the concept of inductive risk" (Douglas 2000, 560–1). Of course, the term "inductive risk" was coined only in 1954, so Douglas (2000) is giving an interpretation that what Hempel called "inductive risk" is exactly what Rudner (1953) was pointing to in explaining why science is not value-free. Furthermore, Douglas (2000) is giving an interpretation when she defines inductive risk as "the chance that one will be wrong in accepting (or rejecting) a scientific hypothesis" (561).

As time has gone on, the definition and scope of "inductive risk" have become the subject of debate (Elliott and Richards, 2017). Some contributors, such as Biddle and Kukla (2017, 218) have argued for a relatively narrow definition of inductive risk, as the risk of wrongly accepting or rejecting a hypothesis on the basis of statistical evidence, others, such as Jack Powers (2017), as "the risk of engaging in a scientific practice that is incongruous with the fulfillment of favored criteria based on favored values" (243). To help resolve these debates, we have argued for interpreting inductive risk based on a holistic reading of Hempel's work, and for distinguishing inductive risk from what we call "representational risk" (Harvard and Winsberg 2022).

In our view, a close reading of Hempel's (1935a, 1935b, 1954, 1965, 1981, 2000) suggests the most useful interpretation of "inductive risk": *The risk of endorsing a truth-apt claim whose objective truth value is false*. This interpretation is supported by Hempel's (1935a) explicit recognition that even so-called observation statements (singular statements like "here is a black raven") can never be confirmed, only more or less supported. Thus, Hempel encourages us to recognize that there is inductive risk in accepting not just formal scientific hypotheses (e.g., in the statistical and theoretical contexts that Rudner (1953) and Hempel (1954, 1965) highlighted), but even the most basic of truth-apt claims. Recognizing this prompts us to interpret "inductive risk" broadly enough to encompass all truth-apt claims.

At the same time, we argue, it is important to recognize that scientists do more than accept truth-apt claims. Indeed, when scientists engage in one of their core activities—*building scientific representations, including models*—they are doing something quite different. Philosophers now widely accept that models are not generally evaluated for their truth but for their adequacy for purpose: Building a model does not amount to making a truth-apt claim (Parker 2020). However, that doesn't mean that there is no risk to speak of when it comes to creating scientific representations, e.g., that scientists can't make an ethically significant mistake when building a model. There *is* such a risk, and our values come into play when managing it, just as they come into play when managing inductive risk. We call this risk "representational risk" and we argue that understanding it can help more fully appreciate the role of values in science.

Before we move on, we should note that even a very general concept of "risk" would suffice to show that values play a role in science: If something, *anything*, can go wrong in science, then scientists must use their values to assess the seriousness of that, and to decide how to act in the face of uncertainty. Thus, if our only goal were to show that values play a role in science, period, then the concept of "risk" alone would do. Indeed, Winsberg (2024) has argued that even the concept of *decision-making* alone suffices to show there is an ineliminable role of values in science, since even the most basic, uncontroversial form of decision theory posits that all decisions require the decision-maker to have at least both beliefs (concerning the probable outcomes of our possible actions) and values (concerning the desirability of those outcomes). However, our goal is to give a more nuanced description of when and how values influence science, and to what effect, and for this purpose we argue that it is helpful to distinguish between different *types* of risk that scientists manage. And to make meaningful distinctions, it is helpful to think in terms of *what risk comprises*. One way to do this is to use the "bow-tie" model of risk (Rausand 2013). The point of

the bow-tie model, as we use it, is to separate risk into three parts: (1) An intentional action that involves a possibility of something going wrong (e.g., “riding a bike”); (2) the thing that does go wrong if we are unlucky (e.g., “falling off the bike”); and (3) the consequences that arise from that thing going wrong (e.g., “ripping our pants” and “skinning our knees”). Previously, we have called these three parts of risk the “hazard,” the “hazardous event,” and the “harm,” following [Rausand \(2013\)](#). However, we have since learned that these terms create confusion, particularly since the term “hazard” has different connotations in different disciplines. To prevent this confusion, in this chapter, we will call the three parts of the risk bow-tie the “action,” the “initial undesired outcome” and the “downstream harms.”

Using the bow-tie model of risk, we can separate inductive risk into three distinct parts and contrast it with representational risk.

Inductive Risk

Action: *Endorsing a “fact,” a truth-apt claim about the world with an objective truth value*

Initial Undesired Outcome: *Endorsing a “fact” whose objective truth value is false*

Downstream Harms: *Whatever bad consequences, specified or unspecified, may follow from endorsing the false “fact”*

Representational Risk

Action: *Making a representational decision in science*

Initial Undesired Outcome: *Making a representational decision that is inadequate for purpose, either for our own purpose or for the purpose of other downstream epistemic agents who will use the representational tool*

Downstream Harms: *Whatever bad consequences, specified or unspecified, may follow from making an inadequate representational decision*

In our view, the bow-tie model of risk encourages us to distinguish between different risks *at least* when we see an important difference in the actions that are the root cause of the harm. Accordingly, one of our key claims is that endorsing a fact is importantly different from making a representational decision in science. What we call “representational decisions” are the decisions we make when building scientific representations like models, i.e., choices about “what to represent” and “how to represent” things in the world ([Harvard et al. 2021](#); [Harvard and Winsberg 2022](#); [Winsberg and Harvard 2024](#)), which, crucially, are not decisions about what is true or false. Rather, representational decisions are decisions about what is *adequate for purpose*, which mean they involve pursuing a specified purpose and determining what will constitute adequacy. Such decisions about adequacy for purpose are value-laden on their face, since they straightforwardly reflect both our pursuits and our intuitions about what counts as “good enough” (which are fundamentally linked to values on a number of definitions of the term ([Elliott and Korf 2024](#))). We argue that neglecting to distinguish between decisions about what is “true” and decisions about what is “adequate for purpose” bypasses key insights from the philosophy of modeling and paints an inadequate picture of science, i.e., as an unbroken string of inferences of a single kind. For us, this is already a sufficient reason to distinguish between inductive risk and representational risk.

Sometimes, we encounter people who acknowledge that there is an obvious difference between endorsing truth-apt claims and making representational decisions in science but still see no reason to distinguish between inductive risk and representational risk. For them, it seems, the concept of risk equates to the concept of downstream harm—and the downstream harm of poor representational

decisions appears to them to be “bad inferences,” no different than inductive risk. In this chapter, we will aim to persuade the reader that this is wrong; that is, representational decisions can lead to *unique downstream harms*, not just bad inferences. To be clear, we don’t deny that representational decisions *can* lead to bad inferences: Indeed, we argue that representational risk and inductive risk are in a close relationship with each other, with scientists’ representational decisions directly raising and lowering the chance they will make an erroneous inference. Thus, representationally risky decisions affect the likelihood of the downstream harms we identify with inductive risk. However, some of the potential downstream harms of representational decisions are just not adequately characterized as simple bad inferences. While one might insist on seeing every one of these potential harms as cashing out in the form of a bad inference, we argue that to insist on this is to insist on simplifying the relevant values in science. We demonstrate this below with a case study.

9.3 A case study in representational risk: Analysis of the “LEAP-SIMPLE” model

In this section, our goal is to take a closer look at representational risk by focusing on a number of specific representational decisions in modeling and their potential downstream harms. To do this, we will use a simplified example of a health economics model, which is based on a real model called LEAP (Lifetime Exposures and Asthma outcomes Projection).¹ LEAP is a microsimulation model and an example of what health economists call “Whole Disease” models. These are particularly comprehensive models that represent entire disease systems alongside potential treatment pathways, their corresponding costs, and expected outcomes, for the purpose of informing health resource allocation decisions (Tappenden et al. 2012). In the LEAP model, the disease of focus is asthma, meaning the ultimate goal would be to model all relevant risk factors for asthma, asthma-related outcomes (e.g., asthma onset, asthma exacerbations, and asthma-related hospitalizations), and asthma prevention and treatment strategies. This is a long-term goal, however, and at the time of writing the real modeling team is working on incorporating one specific asthma risk factor (air quality, measured by the quantity of fine particulate matter or “PM2.5”) and evaluating one specific intervention for asthma (the use of high efficiency particulate air (HEPA) filter machines). In our simplified example—the LEAP-SIMPLE model—we imagine that we are working on the same two modeling tasks and that our proximate goal is to calculate the cost-effectiveness of HEPA filter machines for asthma control in Canada. Table 9.1 lists a variety of representational decisions required to build the LEAP-SIMPLE model. Below, we zoom in on a handful of these decisions and reflect on their character and potential impacts, following the framework described by Harvard et al. (2021) and Winsberg and Harvard (2024).

9.3.1 “What to represent”? The risk of distracting, incomplete, or irrelevant model results

The biggest “what to represent” decision in modeling might seem too obvious to discuss: What phenomenon to model in the first place. Famously, philosophers recognize that choices like this (i.e., the topic to study) are intrinsically value-laden, but they brush them off as “epistemically uninteresting” (see Elliott and McKaughan 2009). We argue that to do this is a mistake. Not only is the choice of “what to model” interestingly value-laden, it has the potential to lead to model results that are not “factually incorrect” per se but harmful in other ways. Among these, we argue, are model results that users perceive as harmfully *distracting, incomplete, or irrelevant*.

Consider, first, the choice to represent HEPA filter machines for asthma control. For one, this choice directs people’s attention toward the relationship between asthma and *air quality*, which many

Table 9.1 Examples of Representational Decisions and Potential Solutions in the LEAP-SIMPLE Model³

<i>Representational Decision</i>	<i>Example Solution</i>
What health intervention(s) should be represented?	Use of HEPA filter machines
What population should be represented?	People in Canada with asthma aged 42 and older
What health outcomes should be represented?	Asthma control level; asthma exacerbations; quality-adjusted life years
What levels of asthma control should be represented?	Uncontrolled asthma; partly controlled asthma; well-controlled asthma
How should asthma control level be represented?	Use population prevalence from BILAS (Best Idealized Local Asthma Study) and model variation in response to in-home and outdoor PM2.5 levels
How should outdoor PM2.5 levels be represented?	Use the BILAM (Best Idealized Local Air Model) to obtain average annual PM2.5 and percentage of PM2.5 attributed to wildfire smoke. Multiply by a “scaling factor” corresponding to the predicted increase in wildfire.
How should indoor PM2.5 levels be represented without HEPA filters?	61% of outdoor PM2.5 level
What economic outcomes should be represented?	Costs of intervention with HEPA filter machines; costs of asthma-related healthcare; incremental cost-effectiveness ratio
What costs of the intervention should be represented?	Cost of HEPA filter machines; cost of HEPA replacement filters; cost of electricity used by HEPA machine
What asthma-related healthcare costs should be represented?	Costs of regular doctor’s visits for asthma; costs of asthma exacerbation
What regular asthma medication costs should be represented?	Cost of inhalers; cost of systemic corticosteroids
How should the cost of asthma medications be represented?	Use costs from database ASMED-IDEAL stratified by asthma control level
How should HEPA filter machines be represented?	Represent effect of machine use on electricity consumption; represent effect of machine use on indoor PM2.5 levels
How should HEPA filter machines’ electricity consumption be represented?	87.6 kWh annually
How should the effect of HEPA filter machines on indoor PM2.5 be represented?	69% ($\pm$ 20%) reduction in assumed indoor PM2.5 compared to no HEPA filter machine
How should HEPA filter machine users’ time at home be represented?	16 hours at home per day
What other population variables should be represented?	Age, sex, income, education, Health Service Delivery Area
How should the cost of a HEPA filter machine be represented?	\$150 one-time cost
How should the cost of a HEPA replacement filter be represented?	\$30 ($\pm$ 20%) every 9 months

(Continued)

Table 9.1 (Continued)

<i>Representational Decision</i>	<i>Example Solution</i>
How should the cost of electricity to run the HEPA filter machine be represented?	\$10 annually
How should the HEPA filter machine lifespan be represented?	5 years
How should the need for HEPA replacement filters be represented?	Every 9 months
How will medication use increase in response to increases in PM2.5?	4% increase in salbutamol use per 10 µg/m ³ increase in PM2.5
How will asthma exacerbations requiring an emergency room visit increase in response to increases in PM2.5?	7% increase per 10 µg/m ³ increase in PM2.5
How will people transition between uncontrolled, partly controlled, and well-controlled asthma?	12% probability per month of moving from one state to another
How much direct healthcare costs will be associated with well-controlled, partly controlled, and uncontrolled asthma, respectively?	\$320, \$400, \$425, respectively
How much direct healthcare costs will be associated with an asthma exacerbation requiring an emergency room visit?	\$575
How much utility will be associated with having uncontrolled, partly controlled, and well-controlled asthma, respectively?	0.70, 0.66, 0.61, respectively
How much disutility will be associated with an asthma exacerbation requiring an emergency room visit?	0.007
What time period should be represented?	5 years
How should costs and benefits be discounted?	1.5% annually

people have linked to climate change. Thus, the choice also draws attention to the impacts of climate change and, in this context, effectively endorses a technological solution to the problem. While these choices surely align with many readers' values, we should still ask the following: Could people with different values see this choice as potentially causing harm? Perhaps, some might argue that the choice to model HEPA filter machines will *detract* attention from the relationship between asthma and other variables, such as breastfeeding, the gut microbiome, and early antibiotic use (see [Adibi, Adibi et al. 2021](#))—and, in doing so, reduce awareness of and demand for other, more “natural” interventions to combat asthma. Furthermore, people whose priorities are to address the root causes of climate change might object to interventions that help people adapt to worsening air quality (here, the age-old debate over the relative benefits of mitigation vs. adaptation rears its head). Thus, some could potentially view the results of the LEAP-SIMPLE model as harmfully *distracting*, that is, functioning to manipulate social narratives to align with the “wrong” priorities.

Consider next another basic “what to represent” decision common to many health economics models: What should the health intervention of interest be compared to? As we’ve imagined it here,

the LEAP-SIMPLE model will represent the use of HEPA filter machines and compare it to “no intervention.” While this is a common comparator in health economics, we should remember that the use of HEPA filter machines could be compared to any number of other interventions to improve air quality, such as programs to reduce traffic-related air pollution, to improve building energy efficiency (and hence indoor air quality), or to control sources of indoor air pollution, such as smoking, cooking, or pets. Indeed, anyone who values these potential interventions would wish that the LEAP-SIMPLE included more information, namely the *comparative* cost-effectiveness of HEPA filter machines relative to other strategies to improve indoor air quality. For them, the results of the LEAP-SIMPLE model would be lamentably *incomplete*.

We might also consider the possibility (however unrealistic) that an important number of people simply detest using HEPA filter machines. Perhaps, for some, such machines are just too noisy, too drafty, too bulky, and too high-maintenance: They would be unwilling to use them under any circumstances. For any such people, the results of the LEAP-SIMPLE model would be not incomplete, *per se*, but simply *irrelevant*.

One possible objection to the idea that model results can be *harmfully* distracting, incomplete, or irrelevant is that there is no intrinsically “bad” way to shift people’s attention, that information will always be incomplete in some way and irrelevant from some people’s perspectives. However, this misses the point: Indeed, the very concept of harm is value-laden and essentially dependent on one’s perspective. Furthermore, we expect that many contextual considerations would persuade most readers to think differently about this. For example, would you feel differently about the LEAP-SIMPLE model if you found out it were sponsored by a manufacturer of HEPA filter machines?² What about if the intervention represented were not HEPA filter machines but a program to, say, socially isolate children susceptible to developing asthma? Or, imagine a scenario where the LEAP-SIMPLE model was being built in a community where it was well-known *what* was causing poor air quality. Say, for the sake of argument, that a high-polluting factory had been recently built and many residents had protested it, declaring it their top priority to shut down its operations (or at least significantly reduce its emissions). In this scenario, if the model describing the impact of using HEPA filter machines were being built to purposely hinder the community’s initiative to close the factory, this would surely be seen as harmful from some perspectives. For recent, real-world case studies of harmful distraction in science and accompanying formal analyses, see Freeborn and O’Connor’s (2025) work on “industrial distraction.”

Ultimately, we argue, big picture “what to represent” decisions make a statement about what is a social priority and what is a socially acceptable health intervention, a statement with which people with different values (and beliefs) might reasonably disagree. Even without producing “false” results, there are various ways the LEAP-SIMPLE model could be seen as causing downstream harms, depending on one’s values.

To be sure, we can zoom in to “what to represent” decisions with more granularity than the big-picture research topic. However, these finer-grained decisions are just more detailed articulations of the research question (Winsberg and Harvard 2024). When it comes to asthma, for example, scientists think a huge number of variables might influence the disease onset and so could be considered for inclusion in a model (see Adibi et al. 2021). When variables are not included, it means that understanding their influence on outcomes is just not part of the research question. This is routine in modeling practice, but it still has consequences. For one, it means the model will not be adequate for making inferences about the effects of excluded variables—to do so would increase inductive risk. But on top of that, we argue, it means the model will send a “signal” about what variables are important (and not important) to understand in a given system. Now, of course, sometimes, a variable is left out of a model for good reason, for example, because there are no good data on it. But we should remember that this doesn’t take away the ethical significance of the exclusion, because it doesn’t

change the implications. Furthermore, we should remember that modelers have flexibility: They can always choose to model a variable using poor-quality data or to decline to model altogether because they think a model would be useless without including a certain variable (Harvard et al. 2021).

9.3.2 “How to represent”? Signaling and cascade effects in modeling

In health economics models, what we call “how to represent” decisions often have to do with selecting appropriate data sources to populate parameters and deciding how extensive the sensitivity analyses around those parameters will be. In the LEAP-SIMPLE model, there are two particularly influential choices like this: First, the effectiveness of HEPA filter machines in reducing PM2.5; second, the PM2.5 “concentration response functions” or, in other words, the relationships between PM2.5 and adverse outcomes (see Table 9.1 for the values that correspond to increases in asthma medication dispensations). For both of these values, there is likely to be a huge range of uncertainty, because the values are likely to have been estimated in studies of different designs, in different geographical settings, using different categorization schemes, and so on. And these choices of parameter values have significant, predictable effects on model results, which means modelers will have to judge the acceptability of representing HEPA filter machines as more effective or less effective (by various degrees) than they really are. Furthermore, modelers will have to judge the acceptability of failing to build an informative model, since the wider the sensitivity analyses, the less useful the model generally is to decision-makers. Beyond that, whether or not modelers are aware of it, choosing a certain value for HEPA filter effectiveness or a certain concentration response function has the potential to signal that these are the “true” values or the ones that are adequate for any future model. Indeed, in qualitative interviews focused on understanding the role of values in health economics modeling, modelers have acknowledged that certain ways of parameterizing models often get “entrenched” through a sort of “cascade” effect. The idea is that a certain published model uses a particular value and this sets a precedent (the “signaling” effect). This means using this value becomes a standard accepted practice: Using this value again in a subsequent study requires only citing the published model that used it (the “cascade” effect), whereas using a value that has not been used in a published model would require a strong argument. In their own words, this is how one modeler described the cascade effect in modeling:

[M]e and another guy made it up in a week. That and the assumptions, literally in a week. A big week, because we were on a deadline ... I still see people today using those same assumptions that we made in a week, today, in like [distant country]. Are they the right assumptions? I don't know. They seem alright, but like, you see how that cascade happens.

(Harvard, Werker, and Silva 2020, 7)

Notice that to interpret the risk here as one of “making an erroneous inference,” one would have to work hard to ignore how this modeler describes the situation and the model inputs in question (i.e., as ones they made up in a week and “seem alright”). We argue that this quote is good evidence that modelers generally regard model inputs as needing to be adequate-for-purpose, not true, and the inductive risk lens is insufficient to capture the significance of their decision, including potential signaling and cascade effects in modeling.

Let's look at one more example from the LEAP-SIMPLE model that underlines the potential for signaling and cascade effects in modeling. As we've described it, the LEAP-SIMPLE model is focused on air quality in Canada, where air quality is highly sensitive to wildfire smoke. Thus, to build the LEAP-SIMPLE model, modelers will have to choose how to represent the intensity of future wildfires, given the effects of climate change (see Table 9.1, “How should outdoor PM2.5 levels be represented?”). Presumably, all other things being equal, we can expect that wildfires will increase

as temperatures increase, but climate scientists have warned us that current climate models aren't good at predicting wildfires ([Sanderson and Fisher 2020](#)). This is because wildfires result from complex interactions between a number of variables beyond temperature. (Some variables that reduce wildfire, like precipitation, could even go up because of climate change. Furthermore, increasing wildfires will eventually lead to loss of "fuel," which will eventually lead to fewer wildfires. It's all a very complex non-linear system with many feedbacks, both positive and negative.) But among the variables that affect wildfire are human activities (beyond just global emissions of greenhouse gases), which are typically left out of models. Environmental modeling studies can project wildfires under a variety of greenhouse gas emissions scenarios, but there will remain significant uncertainty around their results. The bottom line is that choosing how to represent future wildfires in a health economic model is a highly unconstrained decision (see [Harvard and Winsberg 2024](#)).

In this context, we should recall the signaling potential of models. One thing to know is it's currently very common to emphasize that wildfires will surely increase due to climate change, with many sources appearing to imply a preference not to under-estimate future wildfire severity (e.g., [The Global Climate and Health Alliance 2021](#); [United Nations Environment Program 2022](#)). When it comes to choosing how to represent future wildfires in health economic models, there are ethical considerations that suggest modelers should over-estimate rather than under-estimate. But we might ask the following: Could representing wildfire as necessarily intensifying have a negative impact? Perhaps it could, if this signals to policymakers that wildfires are beyond the influence of local policies when this might not be true and causes harmful cascade effects, including ignoring or dismissing potentially effective initiatives to control wildfire.

On this topic, one known possibility is to use "prescribed burns" to help control wildfires ([D'Evelyn, Jung, and Alvarado 2022](#)). But these practices face regulatory hurdles due to health concerns, since prescribed fires themselves create smoke. However, the comparative effects of prescribed burns versus natural wildfires haven't been conclusively established. It is, in fact, a hotly contested question that often arises in local debates about the use of prescribed burns. Ideally, we would like to have the best information about this, so that we could optimize the use of prescribed burns (i.e., maximize their benefits while minimizing their harms). In this context, we might worry that representing wildfire as necessarily intensifying, rather than uncertain, could undermine this goal. That is, representational choices could signal to decision-makers that interventions to control wildfire don't exist at all, when they do, or that they're on balance harmful, when this is unknown. Notice that this concern is in tension with other concerns that modelers may have, e.g., about giving the impression that climate change isn't serious and that wildfire isn't a major threat. Decisions about how to model future air quality, therefore, are highly value-laden decisions. But they aren't decisions that, for the modeler as a decision-maker, are about what truth-apt claims to endorse or not endorse. They might not even be just decisions about what representational choices are most adequate for the purpose of making the most accurate inferences. They could be decisions about which representational choices promote the best messaging to support what modelers take to be the public good.

9.4 Conclusions: Future directions for philosophers of values in science

In this chapter, we have tried to persuade the reader to distinguish between inductive risk and representational risk. For us, embracing the conceptual distinction between inductive risk and representational risk (as we have defined it) has many pay-offs: It invites us to look again at the so-called epistemically uninteresting choices in science (see [Elliott and McKaughan 2009](#)) and it points us to our patterns of attention and inattention in scientific inquiry, asking "what *exactly* should we worry about here?". As we have argued, we think that drawing factually wrong conclusions is just one important thing to worry about in science. We should be no less worried about building scientific

representations that may get some things right, but at the same time actively distract us, steering us away from other, more significant truths.

In the future, we hope that more philosophers will come to see representational risk as a fruitful lens through which to analyze the role of values in science. Indeed, there are many future questions that philosophers might ask about representational risk. For example, does representational risk come in different *types*? In a recent paper, Quill Kukla (2024) has argued that it does, using the example of map-making to highlight differences between what they call “aesthetic risk,” “categorization risk,” and “simplification risk.” This is a fruitful analysis, which encourages us to think deeply about the value-laden decisions that are necessary in map-making. At the same time, philosophers may want to question the apparent overlap between Kukla’s three types of representational risk. One might ask, for example, is there not a relationship between simplification risk and categorization risk? After all, some consider the use of too-few categories to be a harmful oversimplification in some contexts. And what about the relationship between simplification and aesthetic outcomes? In our view, one way to potentially strengthen future analyses inspired by Kukla’s (2024) would be to distinguish between epistemic risks more systematically, on the basis of significant differences in the *action*, *initial undesired outcome*, and/or the *downstream harms*. At least, we think distinguishing between “types” of representational risk would make sense if we can successfully argue for significant differences between representational decisions. At the same time, the commonalities between representational decisions may be more important than the differences between them. We hope that future work on these questions will foster a more nuanced understanding of values in science and modeling.

Notes

- 1 We thank the modeling team for allowing us to use the LEAP model as the basis for our simplified example. We should emphasize that we have changed many details in order to make the example suitable for this short chapter. None of what we present here reflects the team’s real/final modeling decisions. For readers interested in the LEAP project, please see Adibi et al. (2024), which reports the team’s early work on the topic and also informed the simplified examples used in this chapter.
- 2 This is a good place to remind the reader that the LEAP-SIMPLE model isn’t a real model (and the model that inspired it is not sponsored by a HEPA filter machine manufacturer).
- 3 This table is modified from Adibi et al.’s (2024) Table 2 for illustration purposes. It should not be considered a source of scientific information.

References

- Adibi, Amin, Prabjit Barn, Erin M. Shellington, Stephanie Harvard, Kate M. Johnson, and Christopher Carlsten. 2024. “High-Efficiency Particulate Air Filters for Preventing Wildfire-Related Asthma Complications: A Cost-Effectiveness Study.” *American Journal of Respiratory and Critical Care Medicine* 209 (2): 175–84. <https://doi.org/10.1164/rccm.202307-1205OC>.
- Adibi, Amin, Stuart E. Turvey, Tae Yoon Lee, et al. 2021. “Development of a Conceptual Model of Childhood Asthma to Inform Asthma Prevention Policies.” *BMJ Open Respiratory Research* 8 (1): e000881. <https://doi.org/10.1136/bmjresp-2021-000881>.
- Biddle, Justin B., and Rebecca Kukla. 2017. “The Geography of Epistemic Risk.” In *Exploring Inductive Risk: Case Studies in Values in Science*, edited by Kevin Elliott and Ted Richards. Oxford University Press.
- D’Evelyn, Savannah M., Jihoon Jung, Erneston Alvarado et al. 2022. “Wildfire, Smoke Exposure, Human Health, and Environmental Justice Need to Be Integrated into Forest Restoration and Management.” *Current Environmental Health Reports* 9: 366–85. <https://doi.org/10.1007/s40572-022-00355-7>.
- Douglas, Heather. 2000. “Inductive Risk and Values in Science.” *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. “How Values in Scientific Discovery and Pursuit Alter Theory Appraisal.” *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Elliott, Kevin C., and Rebecca Korf. 2024. “Values in Science: What Are Values, Anyway?” *European Journal for Philosophy of Science* 14: 1–24. <https://doi.org/10.1007/s13194-024-00615-3>.

- Elliott, Kevin C., and Ted Richards, eds. 2017. *Exploring Inductive Risk: Case Studies of Values in Science*. Oxford University Press.
- Harvard, Stephanie, and Eric B. Winsberg. 2024. "Managing Values' in Health Economics Modelling: Philosophical and Practical Considerations." *Social Science & Medicine* 358: 117256.
- Harvard, Stephanie, and Eric Winsberg. 2022. "The Epistemic Risk in Representation." *Kennedy Institute of Ethics Journal* 32 (1): 1–31.
- Harvard, Stephanie, Eric Winsberg, John Symons, and Amin Adibi. 2021. "Value Judgments in a COVID-19 Vaccination Model: A Case Study in the Need for Public Involvement in Health-Oriented Modelling." *Social Science & Medicine* 286: 114323. <https://doi.org/10.1016/j.socscimed.2021.114323>.
- Harvard, Stephanie, Gregory R. Werker, and Diego S. Silva. 2020. "Social, Ethical, and Other Value Judgments in Health Economics Modelling." *Social Science & Medicine* 253: 112975. <https://doi.org/10.1016/j.socscimed.2020.112975>.
- Hempel, Carl G. 1935a. "On the Logical Positivists' Theory of Truth." *Analysis* 2 (4): 49–59. <https://doi.org/10.2307/3326781>.
- Hempel, Carl G. 1935b. "Some Remarks on 'Facts' and Propositions." *Analysis* 2 (6): 93–96. <https://doi.org/10.2307/3326609>.
- Hempel, Carl G. 1954. "A Logical Appraisal of Operationism." *The Scientific Monthly* 79 (4): 215–20. <https://www.jstor.org/stable/21072>.
- Hempel, Carl G. 1965. "Science and Human Values." In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, 81–96. The Free Press.
- Hempel, Carl G. 1981. "Turns in the Evolution of the Problem of Induction." *Synthese* 46 (3): 389–404. <https://www.jstor.org/stable/20115601>.
- Hempel, Carl G. 2000. "The Problem of Truth." In *Selected Philosophical Essays*, edited by Carl Gustav Hempel and Richard Jeffrey, 35–74. Cambridge University Press.
- Kukla, Quill. 2024. "Maps and the Epistemic Risks of Visual Representation." *Journal of Language, Mind and the Arts* 5 (1): 39–60. <https://doi.org/10.30687/Jolma/2723-9640/2024/01/002>.
- O'Connor, Cailin, and David Peter Wallis Freeborn. 2025. "Industrial Distraction." *Philosophy of Science* 16: 1–22. <https://doi.org/10.1017/psa.2025.1>.
- Parker, Wendy S. 2020. "Model Evaluation: An Adequacy-for-Purpose View." *Philosophy of Science* 87 (3): 457–77. <https://doi.org/10.1086/708691>.
- Powers, Jack. 2017. "The Inductive Risk of 'Demasculinization'." In *Exploring Inductive Risk: Case Studies in Values in Science*, edited by Kevin Elliott and Ted Richards. Oxford University Press.
- Rausand, Marvin. 2013. *Risk Assessment: Theory, Methods, and Applications*. John Wiley & Sons.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgements." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Sanderson, Benjamin M., and Rosie A. Fisher. 2020. "A Fiery Wake-Up Call for Climate Science." *Nature Climate Change* 10: 175–77. <https://doi.org/10.1038/s41558-020-0707-2>.
- Tappenden, Paul, Jim Chilcott, Alan Brennan, Hazel Squires, and Matthew Stevenson. 2012. "Whole Disease Modeling to Inform Resource Allocation Decisions in Cancer: A Methodological Framework." *Value in Health* 15 (8): 1127–36. <https://doi.org/10.1016/j.jval.2012.07.008>.
- The Global Climate and Health Alliance. 2021. "The Limits of Liveability." <https://climateandhealthalliance.org/bushfires-report/>. Accessed December 28, 2024.
- United Nations Environment Program. 2022. "Number of Wildfires to Rise by 50% by 2100 and Governments Are Not Prepared, Experts Warn." <https://www.unep.org/news-and-stories/press-release/number-wildfires-rise-50-cent-2100-and-governments-are-not-prepared>. Accessed December 28, 2024.
- Winsberg, Eric. 2024. "Managing Values in Science: A Return to Decision Theory." *Kennedy Institute of Ethics Journal* 34 (4): 389–418.
- Winsberg, Eric, and Stephanie Harvard. 2024. *Scientific Models and Decision Making*. Cambridge University Press.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART II

Managing Values in Science



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

10

REFLECTING ON RESPONSES TO THE NEW DEMARCATION PROBLEM

Bennett Holman and T.Y. Branch

The New Demarcation Problem is not a new issue. The fact that [Holman and Wilholt's \(2022\)](#) paper includes a typology of past attempts to give a principled answer to what determines whether a value influence in science is legitimate or illegitimate underscores that this issue had received a good deal of prior attention. Indeed, Dan [Hicks \(2014\)](#) provided a summary of the significant amount of work that had already been done on the topic eight years earlier and argued that demarcation concerns required “a new direction for science and values.” Accordingly, one might well ask, what is new in the New Demarcation Problem?

Our first task in this chapter will be to address that question. In this section we begin by providing both a brief summary of Holman and Wilholt's paper and a running commentary regarding the import of that paper to the discourse. Accordingly, even if you are familiar with the paper, this section still aims to offer the reader insight into aspects of the paper that might have been missed. With this summary in hand, the remainder of the chapter will examine the responses to the problem. [Section 10.2](#) examines attempts to solve the problem and argues that they make progress but remain incomplete. [Section 10.3](#) considers arguments that the New Demarcation Problem might need to be reformulated, including the claim that it comprises two separate concerns (ensuring epistemic integrity and maintaining value alignment) and that the authority of science is derivative of satisfying these concerns. Finally, [Section 10.4](#) highlights some proposed extensions of the problem.

10.1 Summary and commentary of the New Demarcation Problem

The problem begins from two central claims. The first is that [Holman and Wilholt \(2022\)](#) take for granted that the value-free ideal (VFI) is a non-starter. While there seems to be no end to the new defenders of the ideal, such defenders are often advancing objections that have already been dealt with ([Brown 2024](#)). Thus, the first starting point recognizes that within the group of professional philosophers that have an expert command of the topic, there is a near consensus that values should have some influence on the core of science (e.g., which assumptions to accept, methodological choices, interpreting data, etc.). The second claim is a caution against opening the floodgates to every value influence. For example, it is clear that one cannot simply fabricate data and justify it on the basis of one's values.

From these two claims it follows that some principled standard must be given for classifying some value influences as legitimate and others as illegitimate; “just where and how this line is to be

drawn is ... the New Demarcation Problem” (211). But as we said above, this question had already been posed, so what are the contributions of the paper itself? In this section, we will summarize the four central contributions of the paper: (1) A framing of the issue as a “demarcation problem”; (2) a typology for ordering previous answers; (3) a novel case study significantly different from others in the literature; and (4) a standard for a solution to the problem.

10.1.1 The demarcation framing

As many readers will no doubt be aware, the title of the paper is an allusion to Karl Popper’s demarcation problem. Living in early twentieth-century Vienna, Popper was surrounded by revolutionary theories in physics (Einstein’s theory of relativity); psychology (Freudian psychoanalysis); and history (Marx’s dialectical materialism). Yet it seemed to Popper that relativity had something that the other two theories lacked. To be clear, it was not that he thought that only relativity was true, indeed at the time it seemed to him very likely that Einstein was wrong. Rather, Popper’s central contention was that relativity was scientific, whereas psychoanalysis and dialectic materialism only had pretensions to be science. Popper’s demarcation problem was to come up with a principled answer that bore out these judgments and that could be used to demarcate science from pseudoscience.

Hanyu Ma (2023) highlights a strong connection between the old and new demarcation problems, noting that they share the epistemic aim of generating reliable knowledge. Importantly, acknowledging this connection is not tantamount to suggesting that the result of illegitimate value influences is pseudoscience. Illegitimate value influence produces science, but science that cannot fulfill a function it traditionally serves in a democratic society. Accordingly, I (BH) have found it useful to refer to such science as *deficient* (Holman Unpublished).

A further benefit of seeing the legitimate/illegitimate values debate as a demarcation problem is the possibility that the current discourse might benefit from debates around the Popperian demarcation problem. For example, both Koskinen and Rolin (2022) and Resnik and Elliott (2023) note that an important point in the Popperian demarcation debate was an abandonment of developing necessary and sufficient conditions that cleanly separate science from pseudoscience. They argue that the New Demarcation Problem will also lack a crisp solution.

10.1.2 Categorization

A second contribution is a categorization of the previously proposed solutions. Fundamentally, the typology is organized around the presupposition that

it is inevitable that (1) certain non-epistemic values are employed (2) in at least some roles within the research process in ways that (3) have consequences for the generation of knowledge. These values are (4) more or less well coordinated with expectations within and without the research community, and they (5) take place within a broader context that is more or less well-suited to facilitate critical discussion of them.

(2022, 212)

Each of these numbered features provides an avenue of some attempted solution.

A fault line in this typology is whether the content of the values matters. In the first instance, (1) *axiological* solutions argue that certain values are the right values, and the use of these values is legitimate. Of course, just because certain values are incorporated into science does not necessarily entail that the outcome will further those values. Accordingly, (3) *consequentialist* strategies legitimate the influence of values not based on what values are used, but whether the use of those values

promotes the right ends. In contrast, (2) *functionalist* accounts are completely agnostic about which values are employed, so long as those values play the right role in scientific reasoning. Similarly, (4) *coordinative* solutions are agnostic to which values are used, so long as they are properly aligned with the norms and expectations of the proper reference group (e.g., the public, stakeholders, other scientists, etc.). Finally, there are (5) *systemic* solutions that focus on the scientific community. Such accounts suppose that if the scientific community has the right social structure then the result will be a legitimate use of values.

Importantly, [Holman and Wilholt \(2022\)](#) note that many actual accounts combine multiple ideal types. For example, Heather [Douglas \(2009\)](#) is well-known for arguing in favor of allowing values to affect decisions like how to code ambiguous data. However, she has also argued that the values incorporated into scientific research should reflect the values of the public ([Douglas 2005](#)). Accordingly, Douglas' account is both a *functionalist* and a *coordinative* solution. To see more detailed examples and how prominent views are categorized, the interested reader should refer to the original paper.

10.1.3 Case study

The case study at the heart of the article revolves around an irrigation canal in the Upper Klamath Lake in the Cascade mountains just north of the border between Oregon and California. For years the US Fish and Wildlife Service (USFWS) had monitored the health of two endangered species of suckers, fish that were also central to lives of local indigenous peoples and which the federal government had an obligation to protect. A main focus of concern for the USFWS was the height of the lake, in part, because it was an aspect of the environment that could be controlled. Via the Link River Dam, water from Klamath Lake can service over 1,000 small farms, but only by lowering lake water levels.

A flash point was reached in 2001, when the USFWS recommended that the minimum acceptable water level be raised and water to the farms cut off. This recommendation was based upon mechanistic reasoning about the local ecology (e.g., higher water levels reduce the odds of mass die offs from anaerobic conditions) and came from an organization who was tasked with protecting the endangered species. In response to protests from the farming community, the Bush administration tasked the National Research Council (NRC) with reviewing the work of the USFWS. After completing an analysis of critical events (e.g., mass die offs) and lake elevation, the NRC concluded that there was no clear relationship between lake elevation and the health of the sucker population. Unsurprisingly, this analysis was heralded by the farmers.

The Klamath case is very different from most, if not all, of the case studies in the values and science literature. Because the focus was previously the VFI, philosophers selected cases they hoped would make their judgment intuitive to the reader (i.e., cases where the use of values influences seemed obviously legitimate (or obviously illegitimate)). For example, [Anderson \(2004\)](#) recounts previous research into divorce which claimed that women were hurt financially because divorce led to lower incomes. In contrast, researchers influenced by feminist values interviewed women and found that many of them experienced their financial situation as having improved because while their income had decreased, they often had greater fiscal control. The use of values seems unobjectionable and consonant with better science. We think of these as easy cases.

The value influences at Klamath are more complicated. Indeed, some early reactions have suggested that the purpose of the case study is unclear, so here we provide further clarification. The intention behind selecting Klamath was to find a case where it wasn't clear whether the value influences were legitimate. Rather than a conflict between corporate interests and public good, or a case where the use of values clearly leads to better epistemic results, Klamath presents a case where various different publics (and various different private interests) have conflicting values. In Klamath, different sets of values are furthered by the USFWS and the NRC reports but pinpointing how exactly values

shaped the research is less clear. Neither the USFWS nor the NRC was manifestly epistemically suspect, yet each side in the public debate discredited the report that cut against their interest by pointing to the values that the disfavored study promoted.

That parties with contradictory views could each claim to be supported by the best science runs contrary to American statesman Daniel Moynihan's (1983) famous quip that, "everyone is entitled to [their] own opinions, but not [their] own facts." In part, this is because political debate about how to manage tradeoffs ideally begins with a shared agreement about what the tradeoffs are (John 2021). As seen in Klamath, it would appear that once scientific results become value-dependent, they can no longer serve this function (Holman 2021, Unpublished). However, if there were a replacement to the VFI that allowed for some value influences, but not others, we may yet find that value-laden science can still establish a shared body of facts (Holman Unpublished; Schroeder 2021).

Again, Klamath was chosen because it is more representative of the ways in which values inform science. Ideally, a solution clarifies the situation, either by giving a principled reason for privileging one analysis over the other, or by explaining why there is no such simple answer. Whatever the resolution, this case highlights the challenges encountered when using any particular set of values in the face of a pluralistic and polarized public. Hopefully, a successful solution to the New Demarcation Problem would be able to take such cases and to give a clear answer as to how values can legitimately influence research without preventing science from serving the crucial functions that we wish it to play in a democratic society.

10.1.4 Chesterton's fence

The final philosophical contribution of the paper is that it offers a set of desiderata that any successful replacement to the VFI must possess. The grounding of this claim borrows from the insights of twentieth-century social critic G.K. Chesterton. In this section, we will recount the parable and then summarize how it applies to the values in science debate.

In the parable, one reformer points to a social institution, a fence for simplicity, and says to the second, "This fence is useless, may I tear it down?" To which the second replies, "If you see no purpose for this fence, then go away until you do see one, and only then might I permit you to tear it down." At first blush, the response is paradoxical. Why should one have to provide a reason to keep the fence, if one wishes to destroy it? The wisdom encapsulated in Chesterton's fence is that social institutions are purposive creations. Absent knowledge of those purposes, one cannot know if their proposed reform is an improvement over the status quo. Indeed, the benefits of a well-functioning social institution may be obscured. It may be much easier to see that a fence impedes one's walk each day than to realize that it keeps out the wolves each night.

Thus, a rational proposal for a change begins with an account of the original purposes. Next, if the reformer accepts the purposes, they must provide an account of how the reform will better serve those ends. If they do not accept the original purposes, they must provide an account of why those purposes were bad, have become bad, or why they are purposes which are no longer being served. If the original purposes are bad, the reformer must also provide replacements.

Turning to the VFI, we find that three purposes were offered by its original defenders: Veracity, universality, and authority. That is, scientists should aim to produce truth (veracity), to generate knowledge that can be used freely for purposes not intended by the producer (universality), and to produce trustworthy knowledge that can establish a shared body of facts for social discourse (authority). The final claim is that none of the proposed replacements for the VFI successfully serves all of these purposes. Perhaps these are bad purposes, have become bad purposes, or are purposes which are no longer being served, but that must be shown by the would-be reformer. Accordingly, if one follows Chesterton's logic, we have no satisfactory alternative demarcation of legitimate from

illegitimate value influences and thus, no satisfactory alternative to the VFI, which is not to say that none is forthcoming but only to say that the New Demarcation Problem is an open question.

10.2 Responses and rejoinders

The interest stimulated by the paper has already led to significant progress in the debate. In this section, we will survey responses to the problem organized by the typology discussed in 10.1.2. In each case, we try to stress the strengths and the shortcomings of these solutions in hopes of stimulating more compelling solutions.

10.2.1 Functionalist responses

Some philosophers in the values in science debate have responded to arguments against the VFI with alternatives that admit only minor modifications to the VFI (John 2021; Sung Unpublished). Their motivations differ, but the strength of these accounts is that they are responsive to some of the concerns that motivated the adoption of the VFI in the first place. Specifically, they focus our attention on the need to preserve the veracity (Steel 2017) and the authority (John 2021) of science.

Stephen John (2021) in particular has proffered a “truth-first” response to the New Demarcation Problem. He begins from the fact that members of the public generally cannot assure themselves of a scientific claim by direct engagement in the science; they must take it (or leave it) based on trust. In order to trust in value-laden science, a member of the public requires two things. They must know that the epistemic standard that the relevant community upholds is sufficient to compel their assent. And they must have social assurance that research in question adheres to that standard. However, because scientific research must be able to achieve these ends in the face of value pluralism, scientists must adopt “high epistemic standards” that are generally acceptable, irrespective of one’s idiosyncratic values.

A similar argument has been put forward by Doohyun Sung (Unpublished) in his consideration of method validation in toxicological studies. Sung examines how toxicological labs introduce new methodological standards and make decisions about when to consider a method validated. He draws on Steel’s (2017) account, which allows the use of non-epistemic values only when they do not impinge on epistemic values. In the context of method validation, Sung argues that coherence between old and new methods plays this role in method validation. For example, while researchers may see the ethical concerns entailed by reliance on animal testing, they should not relax epistemic standards ensuring coherence (the new method serves the same ends as the former) simply because the newer methods are ethically superior. If we give in to accepting animal-free methods on flimsy epistemic grounds, we may seriously impair future inquiry and/or be more likely to allow dangerous products to be approved and widely used.

Finally, Philippe Stamenkovic (2023) has argued that Sven Ove Hansson’s corpus model (as a functionalist strategy) is the best model going to address the New Demarcation Problem. Hansson’s corpus, or the total body of accepted and consensual scientific knowledge published in the literature, is restricted to claims that meet the highest evidential levels required for all possible applications of the claim. Once such a standard is met, knowledge can be used for any application, theoretical or practical (universal). The extended process of scientific debate preserves or even strengthens the epistemic integrity of scientific knowledge (veracity), and the knowledge produced as a result is thus trustworthy (authority).

Two problems arise with such accounts. The first is that it severely restricts the role of values to either have minimal impact on scientific reasoning or severely restricts the content of values to only those which are widely held. Perhaps this limited role for values is as far as one can go without

causing more problems than one solves; however, many of the arguments against the VFI also apply to allowing only a minimal role for values. Accordingly, most philosophers advocate for a far more expansive role.

A second issue is that it discounts the consequences of false negatives. Suppose that scientists only claim that an effect occurs when almost all parties will accept such a judgment. In such cases, science will have authority. However, we often seem to have grounds for acting on scientific claims before such high evidential thresholds are met. For example, in the Klamath case study, one plausible interpretation of the case is that the NRC was applying such high standards. Now, if the NRC employed high standards and claimed that the evidence conclusively showed that the water level should be raised, then the general public could trust that result. However, if, as occurred in Klamath, the NRC declared that the evidence was insufficient then this declaration should not be treated as authoritative by all parties because given a high value for protecting the endangered suckers, there was sufficient evidence. If our concern was avoiding false positives, such an approach may have merit, but if we wish to balance false positives and false negatives the justification for such approaches falls apart (Wilholt 2022).¹

10.2.2 Axiological approaches

The early proponents of the VFI wanted science to serve as a “transpolitical authority”. That is, a social institution “that is [generally] acknowledged... as adjudicating political debates and/or establishing a shared body of fact” (Holman 2020, 371). Whereas functionalist proposals have largely focused on epistemic considerations as a means of grounding the authority of science, axiological solutions seek to privilege one set of values over others.

For example, Chris ChoGlueck has argued that equity should guide FDA trials. His wonderfully detailed case study explores how individualistic standards of assessing risks in drug trials have prevented male birth control from becoming commercially available. While male and female birth control have similar risk profiles, these risks are judged to be acceptable for women because it prevents them from becoming pregnant and being subject to the resulting risks (e.g., deep vein thrombosis). Male birth control has the same benefits, but since they don’t accrue to the person taking the birth control, they are not factored into the risk-benefit analysis in evaluating safety. ChoGlueck argues that equity considerations motivate analyzing the risk/benefit profile of the procreating couple, by which standard both male and female birth control would be approved.

Some axiological approaches are also systemic approaches in that they specify a procedure to pick out the right values. For example, Frank Cabrera’s (2022) “Rawlsian Solution” stipulates that the right values are the values that would be selected from behind the veil of ignorance according to the principles of justice. Likewise, Kitcher (2011, 2024) argues that the right values are the ones that would be selected over the course of an ideal deliberation among all stakeholders.

Yet even if scientists could secure superior ethical knowledge, it is not clear that the result of infusing those values into their science would be trusted by the public (Holman Unpublished; Holman and Wilholt 2022, section 5.1). To return to ChoGlueck’s case, we might note that whatever the ethical bona fides of “equity” are as a value, one thing the first few months of the Trump administration has made clear is that “equity” is not generally endorsed in contemporary American society. As Bright (2018) has previously pointed out, it may well be that the VFI is sub-optimal from an epistemic—and we might add an ethical—perspective, but it may be necessary to maintain authority. Alternatively, perhaps we should not want science to be trusted and/or trustworthy, but if that purpose of the VFI is to be replaced, then further argumentation is needed to grasp the nettle.

Despite these difficulties I (BH) do think that axiological/systemic approaches are among the most promising solutions to the New Demarcation Problem, particularly when they incorporate a coordinative mechanism (see also Lusk 2021; Schroeder 2021). The main stumbling block seems to

be overcoming distrust. The public generally trusts scientists' epistemic superiority but balks at the idea that they have some special ethical knowledge. If there is not an objectively correct set of values, then the central tension stems from value pluralism. There are no grounds on which one is rationally obligated to accept a scientific finding that depends on values that one does not share (Geislar and Holman 2025). Yet in many cases (e.g., climate change) there may be good reasons to forge on ahead with unshared values and to declare that certain positions are beyond the bounds of reasonable discourse. For science to resolve this tension, there must be grounds to deem some parts of science authoritative, even though they rely on unshared values. If the grounds of the resolution are not rational, then a political resolution is the most plausible alternative. Nevertheless, no current account is fully satisfactory (Holman Unpublished).

10.2.3 List-based solutions

As was discussed in Section 10.1.1, the discourse on the Popperian demarcation tended to the conclusion that there was not going to be a set of necessary and sufficient conditions that separated science from pseudoscience. Multiple authors have suggested that the same may be true for the New Demarcation Problem and advocated for list-based solutions as a result (Koskinen and Rolin 2022; Resnik and Elliott 2023). This approach is orthogonal to the original typology. It is characterized by providing a list of ideals rather than articulating a single principle of legitimacy.

Koskinen and Rolin (2022) suggest that such a list might be garnered by abstracting the principles of three influential arguments against the VFI, though they allow that other principles can be added should it be necessary. Their three principles are the following: (1) Non-epistemic values can guide the choice of epistemic values, but they cannot supplant epistemic values; (2) non-epistemic values can play a role in scientific reasoning (e.g., setting evidential thresholds); and (3) non-epistemic values can inform selection of background assumptions. In exploring the reasoning behind these principles, Koskinen and Rolin (2022) frequently return to concerns about epistemic integrity of science and about whether risk assessments are based on sound moral judgments. This is a strength of their work, and we will return to it in Section 10.3. Our main issue with this approach is that the principles seem to be in tension. Philosophers that advocate for (2) and (3) typically endorse a more expansive role for values, including in cases that are detrimental to epistemic concerns. Moreover, if (1) is indeed a strictly stronger restriction and one accepts it, then (2) and (3) provide examples of how (1) might be instantiated, but do not constitute additional restrictions.

Another list solution is presented by Resnik and Elliott (2023, 273) who argue that “efforts to distinguish between legitimate and illegitimate influences should focus on whether researchers are complying with epistemic and ethical norms that are constitutive of good science, rather than on some particular criteria of legitimacy.” Consequently, they propose a list of questions that might be asked to determine whether the relevant research is exemplary or deficient (e.g., is the study well-designed to achieve its aims?). Moreover, rather than ask whether the science is good or bad, they argue we should be asking whether it is “good enough” for the proposed purposes.

This provides a helpful articulation to the many authors who have spoken generally about “epistemic integrity” without articulating exactly what that amounts to. For example, Reutlinger (2022) has argued that a non-epistemic value is epistemically illegitimate if and only if a non-epistemic value caused a scientist to make an epistemic error. One of the strengths of Resnik and Elliott's paper is that the paper produces a detailed list of ways to assess whether the epistemic norms of science have been upheld. On the other hand, very little attention is paid to what must be done to ensure authority, and no consideration is given to universality. In line with the reasoning of Chesterton's fence, if authority and universality cease to be desiderata, some further arguments must be provided.

Finally, list solutions are not logically entailed from rejecting a search for the necessary and sufficient conditions of legitimacy. A single principle of legitimacy may be proffered that can be met to varying degrees. The result of such a solution would be a gradation of legitimacy based on a single principle.

10.3 Reframing the New Demarcation Problem

Rather than offering categorizable solutions, some philosophers have argued against the way the issues have been framed. In some cases, this has meant arguing that the New Demarcation Problem is poorly posed. In others, this has entailed arguing that the focus of the current debate is too parochial to offer a complete answer.

An example of critiquing the coherence of the New Demarcation Problem is Wendy [Wagner's \(2022\)](#) proposal that it runs together ends-oriented bias and the professional values problem. The former occurs where values lead scientists to sacrifice the epistemic reliability of science. The latter occurs when the values of the scientific community are not representative of the general public.² Crucially, because the solutions are distinct, the problems must be kept apart. For example, to address the ends-oriented bias problem, Wagner recommends subjecting research to a blue-ribbon panel review of well-qualified and independent scientists. But notice that this solution would leave untouched, if not exacerbate, the professional values problem, which requires opportunities for public input to be addressed.

Inmaculada [de Melo-Martín \(2024\)](#) agrees with Wagner but goes further to suggest that two other concerns raised by philosophers can be subsumed into these problems. First, she argues that the only reason a member of the public might justifiably distrust science is either because of the imposition of an unshared set of values or because the scientists' use of values produced an epistemic distortion. She also asserts, contra some axiological solutions, that there is no set of correct values and so, in principle, any values can be used in science as long as they are properly chosen. Moreover, if values are "selected following procedures that attend to the interests of relevant stakeholders" (859) grounds for distrust no longer exist, even for those who regard the values selected as reprehensible.

The final two points merit further attention. Is it really the case that a person should trust science that incorporates unshared values? Both [John \(2021\)](#) and [Geislar and Holman \(2025\)](#) have argued that, at least on rational grounds, they should not. Might there be a political solution to the demarcation problem that can support de Melo-Martín's claim? Perhaps, but no currently available solution is sufficient ([Holman Unpublished](#)). Further, whether the issues around trust are necessarily derivative of addressing issues of epistemic integrity and value imposition cannot be decided a priori; since a solution to the New Demarcation Problem may be accompanied by a new conception of the purposes of science, whether epistemic degradation and value imposition are the only reasons to object to the use of values in science is a claim that can only be assessed with a specific solution in hand. Furthermore, [de Melo-Martín's \(2024\)](#) focus with respect to authority is only on science being trustworthy, but as discussed in [Section 10.2.3](#), authority entails that it is both trustworthy and trusted. de Melo-Martín provides no argument that we should abandon our aspiration that science actually be trusted, not merely trustworthy according to some philosophical standard. Accordingly, her account here is at best incomplete. Similarly, whether a solution should be agnostic as to which values are used is a long running, though perhaps underexplored disagreement and more direct attention to this split would be fruitful.

In contrast to reformulating the New Demarcation Problem, other philosophers have argued that our philosophical view needs to be expanded to new vistas in order to provide a satisfactory answer. For example, it has previously been shown that the pharmaceutical industry can promote their

interests simply by identifying researchers that already pro-industry values and providing them with resources that allow these researchers to gain prominence in their field (Holman and Bruner 2017). Marcel Hobma (2023) has extended this insight and argued that a satisfactory solution to the New Demarcation Problem requires understanding how values originate, proliferate, and transmit between the public and the institutions and individuals that compose the scientific community. Indeed, attention may even need to move beyond actors to the way that values are embedded in scientific instruments (Karakas and Tuboly 2024).

A more expansive perspective may be particularly crucial if the solution to the New Demarcation Problem also seeks to reimagine the traditional purposes of science (veracity, authority, and universality). Scientific advising, science education, and science communication all presuppose the VFI in the way they interface with society (Branch and Douglas 2024). Accordingly, a drastic departure from the VFI may also require a new social contract for science (Douglas and Branch 2024; Brown 2025).

Finally, focusing on legitimacy, Hannah Hilligardt (2022) rightly points out that social perspectives, opinions, and interests also affect the authority of science. However, while a broader understanding of social legitimacy may be independently interesting, solving the New Demarcation Problem does not require specification of everything that might threaten the veracity, universality, or authority of science. One needs only specify how to uphold these purposes insofar as they are threatened by the infusion of values into science.

10.4 Conclusion

The New Demarcation Problem is currently an extremely lively topic, so much so that this review is almost certainly missing new and important insights published between when it was written and when you are reading. Here we have attempted to summarize and critique only the first wave of responses in hopes that our summary will help the reader see how the New Demarcation Problem frames the debate and that our critiques will stimulate future contributions to push on past the limitations identified here.

In closing, we wish to suggest one further lesson of the previous demarcation debate. Popper (1963) held that, *“genuine philosophical problems are always rooted in urgent problems outside of philosophy, and they die if these roots decay”* (72, italics in original). Arguably, this occurred as Popper’s demarcation problem became a “philosophical” problem. The debate became desiccated and pedantic as philosophers who took up the demarcation problem focused on trivial pseudosciences like astrology in tackling the demarcation problem. In contrast, Popper was gravely concerned about the threat that Marxist doctrines posed and by answering the demarcation problem, he hoped to deny Marxism the prestige of science and help staunch its growing influence.

The New Demarcation Problem is also motivated by concerns outside the field of philosophy. Feminist scholars made important scientific contributions but found their work discredited due to the influences of values on their work. At the same time, the role that science has assumed in modern society has led to scientific engagement by actors that are significantly affected by non-truth-seeking values. Should these influences find their way into inquiry, the VFI gives one clear ground to object; but abandoning it without a suitable replacement threatens to leave us without the resources to object to those who seek to unjustly assume the power and prestige of science. Finally, this debate occurs at a time when science, as well as many other transpolitical institutions, is being undermined by antidemocratic political movements. Should the discourse lose sight of the real-world grounding of the problem, one should expect a concomitant deterioration in the importance of the work produced.

Notes

- 1 Hansson has developed a way of dealing with this objection by allowing for “non-epistemic adjustments” (Stamenkovic 2023), but we do not find the response convincing. First, this approach can only address using values to set acceptance/rejection criteria, whereas most views envision a far more expansive role for values in science. Moreover, even if this restricted role were accepted, what Hansson has done is cordon off a body of knowledge that can be accepted independent of one’s values and called that “science.” However, all of the same issues remain outside this rarified body of knowledge. Suppose we call what is left “schmience.” If we accept Hansson’s “solution,” science may be value-free, but we have a new new demarcation problem for schmience and values.
- 2 But see Haase (2025) for an argument that discipline-specific values may be legitimate independent of coordination with public values.

References

- Anderson, Elizabeth. 2004. “Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce.” *Hypatia* 19 (1): 1–24. <https://www.jstor.org/stable/3810929>.
- Branch, T. Y., and Heather Douglas. 2024. “Rethinking the Conceptual Space for Science in Society after the VFI.” *Philosophy of Science* 91 (5): 1118–27. <https://10.1017/psa.2023.130>.
- Bright, Liam Kofi. 2018. “Du Bois’ Democratic Defence of the Value Free Ideal.” *Synthese* 195 (5): 2227–45. <https://www.jstor.org/stable/48726196>.
- Brown, Matthew J. 2024. “For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science.” *Synthese* 204 (4): 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Brown, Matthew J. 2025. “Trust, Expertise, and Scientific Authority in a Democracy: A Pragmatist Approach.” In *Pragmatism and Philosophy of Science: Historical and Contemporary Perspectives*, edited by Roberto Gronda, Marianne Janack, and Giancarlo Marchetti. Routledge. <https://doi.org/10.4324/9781003402794>.
- Cabrera, Frank. 2022. “A Rawlsian Solution to the New Demarcation Problem.” *Canadian Journal of Philosophy* 52 (8): 810–27. <https://doi.org/10.1017/can.2023.19>.
- de Melo-Martin, Inmaculada. 2024. “Concerns about Contextual Values in Science and the Legitimate/Illegitimate Distinction.” *Philosophy of Science* 91 (4): 851–68. <https://doi.org/10.1017/psa.2024.20>.
- Douglas, Heather. 2005. “Inserting the Public into Science.” In *Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making*, edited by Sabine Maasen and Peter Weingart. Springer Netherlands. https://doi.org/10.1007/1-4020-3754-6_9.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather E., and T. Y. Branch. 2024. “The Social Contract for Science and the Value-Free Ideal.” *Synthese* 203 (2): 40. <https://doi.org/10.1007/s11229-023-04477-9>.
- Geislar, Sally, and Bennett Holman. 2025. “When Is It Rational to Distrust Scientists?” *Social Epistemology*: 1–18. <https://doi.org/10.1080/02691728.2025.2449600>.
- Haase, Michaela. 2025. “The Nexus of Knowledge, Action, and Values, and the Philosophical, Social, and Normative Foundations of Marketing.” *Journal of Macromarketing* 45 (1): 130–45. <https://doi.org/10.1177/02761467241256027>.
- Hicks, Daniel J. 2014. “A New Direction for Science and Values.” *Synthese* 191: 3271–95. <https://doi.org/10.1007/s11229-014-0447-9>.
- Hillgardt, Hannah. 2022. “Looking Beyond Values: The Legitimacy of Social Perspectives, Opinions and Interests in Science.” *European Journal for Philosophy of Science* 12 (4): 58. <https://doi.org/10.1007/s13194-022-00490-w>.
- Hobma, Marcel. 2023. “The Selection of Functional Values and Practices in Industry-Funded Nutrition Science.” Master’s thesis. Utrecht University. <https://studenttheses.uu.nl/handle/20.500.12932/43516>.
- Holman, Bennett. Unpublished. “Solving the New Demarcation Problem Requires a Theory of Politics.”
- Holman, Bennett. 2020. “STS, Post-Truth, and the Rediscovery of Bullshit.” *Engaging Science, Technology, and Society* 6: 370–90. <https://doi.org/10.17351/ests2020.265>.
- Holman, Bennett. 2021. “An Ethical Obligation to Ignore the Unreliable.” *Synthese* 198: 5825–48. <https://www.jstor.org/stable/48692185>.
- Holman, Bennett, and Justin Bruner. 2017. “Experimentation by Industrial Selection.” *Philosophy of Science* 84 (5): 1008–19. <https://doi.org/10.1086/694037>.

- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- John, Stephen. 2021. "Science, Politics and Regulation: The Trust-Based Approach to the Demarcation Problem." *Studies in History and Philosophy of Science Part A* 90: 1–9. <https://doi.org/10.1016/j.shpsa.2021.08.006>.
- Karakas, Alexandra, and Adam T. Tuboly. 2024. "Materializing Values." *Synthese* 204 (1): 5. <https://doi.org/10.1007/s11229-024-04663-3>.
- Kitcher, Philip. 2011. "Science in a Democratic Society." *Poznan Studies in the Philosophy of the Sciences and the Humanities* 101: 95–112. https://doi.org/10.1163/9789401207355_003.
- Kitcher, Philip. 2024. "The Scientist, Qua Scientist, Is an Ethical Agent." *Filozofia* 79 (3): 231–43. <https://doi.org/10.31577/filozofia.2024.79.3.1>.
- Koskinen, Inkeri, and Kristina Rolin. 2022. "Distinguishing between Legitimate and Illegitimate Roles for Values in Transdisciplinary Research." *Studies in History and Philosophy of Science* 91: 191–98. <https://doi.org/10.1016/j.shpsa.2021.12.001>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Ma, Hanyu. 2023. "Revitalizing the Value-Free Ideal of Science: The Temporal Role of Values." PhD diss., University of Pennsylvania. <https://repository.upenn.edu/handle/20.500.14332/59455>.
- Moynihan, Daniel P. 1983. "More Than Social Security Was at Stake." *The Washington Post*, January 18. <https://www.washingtonpost.com/archive/politics/1983/01/18/more-than-social-security-was-at-stake/87951725-8bfb-426d-933a-dbf1ad94f981/>. Accessed February 7, 2025.
- Popper, Karl. 1963. *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge and Keagan Paul Ltd.
- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54 (2): 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Reutlinger, Alexander. 2022. "When do Non-Epistemic Values play an Epistemically Illegitimate role in Science? How to solve one half of the New Demarcation Problem." *Studies in History and Philosophy of Science* 92: 152–61. <https://doi.org/10.1016/j.shpsa.2022.01.018>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Stamenkovic, Philippe. 2023. "Remarks on Hansson's Model of Value-Dependent Scientific Corpus." *Lato Sensu: Revue de la Société de philosophie des sciences* 10 (1): 39–62. <https://doi.org/10.20416/LSRSPS.V10I1.4>.
- Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values and Science*, edited by Kevin Elliott and Daniel Steel. Routledge.
- Sung, Doohyun. Unpublished. "On Epistemic Risks in Research Method Validation."
- Wagner, Wendy E. 2022. "No One Solution to the 'New Demarcation Problem'? A View from the Trenches." *Studies in History and Philosophy of Science* 92: 177–85. <https://doi.org/10.1016/j.shpsa.2022.02.001>.
- Wilholt, Torsten. 2022. "Epistemic Interests and the Objectivity of Inquiry." *Studies in History and Philosophy of Science* 91: 86–93. <https://doi.org/10.1016/j.shpsa.2021.11.009>.

11

AN INTERMEDIATE APPROACH TO VALUE MANAGEMENT

Wendy S. Parker

11.1 Introduction

According to the value-free ideal (VFI), choices made in the “research” stage of investigation, as scientists produce and analyze data and draw conclusions, should be kept free from the influence of contextual values as much as possible; ideally, such choices should be based on epistemic considerations alone. The rationale for this prohibition on value influence seems straightforward: Contextual values, reflecting what someone wishes or desires to be the case, are irrelevant to choices of methodology that are aimed at finding out what *is* the case, and their influence can bias research in significant ways.

Nevertheless, the dominant view among philosophers of science today is that the VFI should be rejected. According to the *inductive risk argument*, it can be appropriate, and perhaps even morally required, for uncertain choices faced in the research stage of investigation—such as how to classify ambiguous data or which of several plausible models to employ—to be informed by judgments of how bad the consequences of different errors in conclusions would be (Douglas 2000, 2009). Likewise, according to the *aims argument*, when the aims of scientific investigation are both epistemic and practical (reflecting contextual values), it can be appropriate for methodological choices to be justified by appeal to those joint aims (Elliott and McKaughan 2014; Intemann 2015; see also Brown 2017). These are just some of the arguments leveled against the VFI.

As a consequence, the focus of philosophical discussion has shifted to the question of what constitutes legitimate or illegitimate value influence in research, recently dubbed “the new demarcation problem” (Holman and Wilholt 2022; Resnik and Elliott 2023). A range of positions have been advocated, from permissive ones that see science as a force for social justice to restrictive ones that deem appeals to contextual values to be legitimate only when epistemic considerations alone are not decisive.

Another emerging thread of discussion, however, considers the possibility of at least a partial rapprochement among defenders and challengers of the VFI (DiMarco and Khalifa 2019; Lusk and Elliott 2022; Parker 2024). Are there approaches to value management that could be accepted by (some) parties on both sides of the debate? This chapter articulates one proposal, the *epistemic projection approach* (EPA) (Parker 2024). EPA recognizes that sometimes there are good reasons to make scientific research responsive to contextual values, but it achieves this responsiveness via the careful formulation of an epistemic research problem in the problem-selection stage of investigation.

A further attractive feature of EPA, independent of its potential for rapprochement, is that it provides practitioners with concrete guidance on value management. Readers primarily interested in this guidance may focus their attention on [Section 11.4](#).

11.2 Foundations of EPA

Important insights underlie the inductive risk and aims arguments: Scientists *should* take due care to avoid errors that would have particularly bad consequences, and sometimes it *does* make sense for the broader aims of scientific investigation to impact how research proceeds. But in their typical formulations, these arguments are too permissive with regard to value influence.

The aims argument allows that, when research has both epistemic and practical aims, methodological choices may be justified by appeal to those joint aims. This seems right in some cases. For instance, since weather forecasts are produced in order to help people plan their activities, it is appropriate that scientists attend to both accuracy and speed when developing their forecast models; it is no good if highly accurate “forecasts” arrive after the weather has passed. The trouble is that the aims argument doesn’t seem to prohibit investigators from attaching much greater importance to practical aims than epistemic ones such that, in some cases, they might be justified in engaging in shoddy science—or even outright fraud—in order to promote those practical aims. This leads [Steel \(2017\)](#) to argue for a “qualified epistemic priority,” according to which choices in research may be justified by appeal to joint epistemic and practical aims, but the research must respect basic constraints on epistemically adequate science: Investigations must be genuinely probative, data should not be manufactured or cherry-picked, conclusions should take account of relevant evidence, and so on.

Qualified epistemic priority is underwritten by the idea that science is first and foremost an epistemic endeavor; if science is to serve some broader practical aims, it should do so by providing epistemic goods, like information and knowledge, that are relevant to the pursuit of those aims, rather than via persuasion, manipulation, etc. ([Steel 2017](#)). Accepting this view of science, we can say even more. Not only should research respect basic constraints on epistemically adequate science but the default criterion for choosing among routes of investigation should be epistemic quality, i.e., the likelihood that the route will achieve the epistemic goals of the research (if they are achievable at all). This default criterion may be overridden for compelling ethical, pragmatic, or contextual reasons, such as respecting the autonomy of human research subjects, or managing finite resources, or meeting procedural or legal requirements for research that will inform regulations. But reducing the epistemic quality of research should be viewed as a cost, even when on balance such a reduction is worth it for other reasons. This is one foundational commitment of EPA.

A second foundational commitment concerns inductive risk. According to the inductive risk argument, contextual values sometimes have a role to play in resolving uncertain methodological choices. For example, in a study investigating whether a chemical is carcinogenic, if the foreseeable consequences of a Type II error (not concluding that the chemical is carcinogenic, when it is carcinogenic) are judged to be particularly bad, then a researcher might choose to treat ambiguous data, such as tissue samples where there are indicators but not definitive evidence that cancer is present, in a way that minimizes the risk of a Type II error; the researcher might choose to classify those ambiguous cases as cancerous (see [Douglas 2000](#)).

But often a better response to uncertain methodological choices like these is to propagate the uncertainty through to the study’s conclusions (for variations on this response, see, e.g., [Betz 2017](#); [Jeffrey 1956](#); [Menon and Stegenga 2023](#)). Thus, in the example just given, the researcher could calculate how exposure to the chemical affects the risk of cancer, assuming first that ambiguous cases are cancer and then assuming that they are not, arriving at a range estimate rather than a single-valued

estimate of the increased risk. Such an approach is better for multiple reasons. It delivers a more accurate picture of researchers' actual epistemic situation vis-à-vis the question of interest and, when findings are intended to inform practical decisions, it allows for those decisions to be made with awareness of extant uncertainty. In addition, when there is value pluralism in society, permitting just one set of values to influence how a research question is addressed can undermine the democratic legitimacy of scientific findings, erode public trust, and lead to an undesirable fragmentation of scientific knowledge (e.g., "conservative science" informed by one set of values and "liberal science" informed by another) (see [Betz 2013](#); [Menon and Stegenga 2023](#); [Schroeder 2021](#)).

Consequently, EPA incorporates a default preference for propagating methodological uncertainty through to conclusions, either formally or using expert judgment, rather than choosing one methodological option over the others on value-related inductive risk grounds. Once again, this default preference can sometimes be overridden, since in some cases uncertainty propagation is impossible or undesirable. Perhaps implementing one methodological option destroys or consumes the object of analysis, or perhaps it would take too long or be too expensive to explore the implications of the study's uncertain methodological choices, or perhaps the results are being produced for a narrow set of users whose decision-making will not benefit from (and might even be confused by) a careful accounting of uncertainty. It is in situations like these that resolving uncertain methodological choices by appeal to inductive risk considerations (alongside pragmatic ones) becomes more attractive.

In summary, two foundational commitments of EPA are as follows. First, the default criterion for choosing among routes of investigation should be epistemic quality, i.e., the likelihood that the route will achieve the epistemic goals of the research. Second, when there is uncertainty about which methodological choice is best for achieving those epistemic goals, there is a default preference for propagating the uncertainty through to conclusions, rather than choosing one methodological option over the others on value-related inductive risk grounds. These defaults may be overridden for compelling ethical, pragmatic, or contextual reasons. But at a minimum the investigation must respect basic constraints on epistemically adequate science. Otherwise, the activity does not really count as *science*.

11.3 EPA: Values in problem selection

EPA is primarily an account of *how* to make research responsive to contextual values—whether expressed in the form of broader practical aims, ethical constraints, or in some other way—once it has been determined that at least some responsiveness is appropriate. Challengers of the VFI argue that value considerations sometimes appropriately serve as justifying reasons for choices in the research stage of investigation. EPA, by contrast, limits consideration of values to the problem-selection stage. More specifically, EPA proposes that, when there are compelling reasons for research to be responsive to contextual values, this responsiveness should occur by *formulating an epistemic research problem that aligns with those values*. The rest of this section unpacks this proposal.

11.3.1 Problem selection

The problem-selection stage of investigation has received comparatively little attention from philosophers interested in values in science. Often it is simply equated with the choice of a question to be answered. Yet research problems clearly can, and in practice often do, have a richer structure. For instance, an investigation may be undertaken to address several research questions, some of which are of higher priority for answering than others. Or it might be clear from the outset that only information produced before a particular practical decision must be made will be useful, so that the research problem to be solved is not merely to answer a given question but to do so within a specified time frame.

EPA understands research problems in a way that leaves room for this richer structure and allows for making it explicit. In particular, it understands research problems to include not only epistemic goals (corresponding to research questions) but also, optionally, constraints and preferences related to how those goals are pursued. *Epistemic goals* specify particular epistemic goods to be obtained in research, such as particular pieces of information or knowledge. For example, an ecologist might have the goal of learning the main ecological roles of three salient species in a novel ecosystem recently discovered on private land. *Constraints* specify features that research undertaken in pursuit of those goals is required to have. For example, the ecologist might have permission to study the novel ecosystem only insofar as the research does not significantly damage the ecosystem's functioning. *Preferences* indicate features that are desirable, but not required, for research to have. It might be preferred, for instance, that research on the novel ecosystem be conducted using the instruments and other resources that the ecologist's research team already possesses, rather than requiring the acquisition of new ones.

A research problem is *epistemic* in character if its constraints and preferences (if any) reference only features of research qua epistemic activity—its methods and data, the types of errors to which it is more or less prone, and so on; the research problem's specification does not reference any broader aims or contextual values. The example of a constraint given above, that research on the novel ecosystem *must not significantly damage the ecosystem's functioning*, fails to meet this criterion. However, a constraint that simply *excludes the use of method X* would count as epistemic in character, even if the underlying reason for excluding the method was that it carried a substantial risk of damaging the ecosystem's functioning. (In case the reader doubts that it is appropriate to consider this an "epistemic" constraint, consider by analogy that a preference for vegan food is a "dietary" preference, regardless of the motivations for it.)

A research problem *aligns with* a broader aim (or a contextual value expressed in some other form) to the extent that research that achieves the problem's epistemic goals, while meeting its constraints and preferences, can be expected to produce results that are relevant for, useful to, or otherwise helpful with respect to pursuit of that broader aim (or the promotion of the contextual value in some other form). Suppose, for example, that the broader aim motivating an investigation is to make infrastructure decisions that will protect a city from deadly flooding from an adjacent river, given concerns that climate change might bring about such flooding. A research problem that aligns with this broader aim might include the epistemic goal of learning the largest plausible changes in the frequency and magnitude of flooding over the next several decades, as well as a preference for risking overestimating these changes, rather than underestimating them, if some inductive risk must be taken on. A research problem including the opposite inductive risk preference would not align with the broader aim, since underestimating the flooding is more likely to lead to infrastructure decisions that inadequately protect the city.

The process of reasoning from a broader aim (or contextual value expressed in some other form) to an epistemic research problem that aligns with that aim/value is at the heart of EPA and is called *epistemic projection* (Parker 2024). It involves identifying (i) one or more plausibly achievable epistemic goals that are relevant to pursuit of the broader aim (or the promotion of the contextual value) and, optionally, (ii) epistemic constraints and/or preferences regarding how those goals are pursued that will help to align the research with the broader aim (or the contextual value expressed in some other way). The terminology of "projection" here is inspired by the process of geometrical projection, which shows what an object "looks like" when collapsed to a reduced set of dimensions, e.g., a three-dimensional object projected onto a two-dimensional plane; the process of epistemic projection reveals what pursuit of a broader aim (or protection/promotion of a contextual value expressed in some other way) can "look like" when limited to choices regarding epistemic goals, constraints, and preferences that jointly constitute a research problem.

To make (ii) manageable, EPA calls for consideration of just four types of epistemic constraint/preference, which typically present the most significant opportunities for aligning research with contextual values:

- a *Epistemic priorities*: When research has more than one epistemic goal, these constraints and preferences indicate which epistemic goals take priority if trade-offs must be made;
- b *Epistemic resources*: These constraints and preferences identify concepts, models, methods, instruments, etc. that should or should not be employed in research;
- c *Inductive risk*: These constraints and preferences indicate the standards of evidence that should be used when drawing conclusions and which type(s) of error in conclusions should be preferentially avoided and in what circumstances;
- d *Form-of-conclusion*: These constraints and preferences indicate a desired (or undesired) format for research findings to take.

Of course, not every research problem identified via epistemic projection will include all four of these. Sometimes, an investigation will have only one epistemic goal, making (a) irrelevant. Other times, only an inductive risk preference will be important. And so on.

Note, however, that even when there are compelling reasons to try to align research with a broader aim or contextual value, the drawbacks of including particular constraints or preferences in a research problem should be considered: Will meeting the constraint or preference significantly reduce the epistemic quality of the research? Will it be very costly in terms of time, funding, or other resources? Even if including the constraint or preference would help to align research with a contextual value, the foreseeable costs—both epistemic and practical—might be significant enough that, on balance, a decision is made not to include it.

Any constraints or preferences that are included must leave room for research that, at a minimum, respects basic constraints on epistemically adequate science; this is necessary for the research problem to count as *scientifically acceptable*. Thus, for example, a research problem cannot include an epistemic resource constraint—say, requiring the use of a particular methodology—such that research that satisfies that constraint will be virtually guaranteed to produce a particular result, regardless of the truth. Ideally, the research problem will leave room for research that not only respects basic constraints on epistemically adequate science but is of excellent epistemic quality.

11.3.2 The research stage

A research problem identified via epistemic projection then constitutes a brief for the research stage of investigation: Researchers should try to achieve the specified epistemic goals, while satisfying the specified epistemic constraints and preferences (see [Carrier 2022](#) for a similar view in the context of science-based policy advice).

Importantly, the process of epistemic projection will have *screened off* the contextual values from this research stage: The research problem will be specified in a way that makes no reference to broader aims or contextual values, and choices in the research stage will be justified as a means of solving that epistemic research problem, not by appeal to any broader aims/contextual values. (For a concrete illustration, see the research problem specification in [Table 11.1](#), in the next section.) Indeed, in principle the research stage of investigation could be undertaken by researchers who are unaware of the aims/values that played a role in the formulation of the research problem.

If the research findings are shared publicly, then researchers should also report which methodological choices in the research stage were made in response to which preferences and constraints specified in the research problem. This transparency requirement is intended to facilitate healthy scrutiny of research practices. EPA does not, however, require that researchers report the broader

aims or values considered during epistemic projection; these broader aims and values are irrelevant to the assessment of the quality of research qua epistemic activity. Of course, reporting such aims or values, as well as the reasoning that was employed in the process of epistemic projection, is not prohibited either, and researchers might often choose to report this information as well.

11.4 Putting EPA into practice

Having introduced the foundations and basic elements of EPA, this section discusses more concretely how EPA can be put into practice, illustrating with an example involving air pollution. There are three main steps.

Step 1. *Determine whether it is appropriate to try to make the research at hand responsive to contextual values.*

As mentioned earlier, EPA is primarily an account of how to make research responsive to contextual values, once it has been determined that some responsiveness is appropriate. It is beyond the scope of this discussion, and probably a fool's errand, to try to provide an exhaustive list of circumstances in which responsiveness to contextual values is appropriate. But a partial list of representative situations is easy to provide.

First, if research is undertaken to *inform a high-stakes practical decision* where one type of error in conclusions can be expected to have serious negative consequences for the public, then typically it will be appropriate to include an inductive risk preference for avoiding that type of error (if methodological uncertainty cannot be propagated through to conclusions). Second, if research is being performed *in a consulting mode*, then often it will be appropriate to make research responsive to the client's values. A third type of situation is when a regulatory agency conducts scientific research in support of its activities but also has a *value-related mandate*, e.g., to protect human health. Yet another situation is when there is *pervasive value-related bias* in an existing body of research; in some cases, this could warrant making research responsive to an alternative set of values, whether as a corrective epistemic measure or to combat harms ensuing from the existing bias (e.g., if a body of research is sexist or racist).

Example. Suppose a county government is considering granting a permit to a company to build a chemical manufacturing facility in the county. Residents in a city near the proposed site are concerned that the air pollutants that this type of facility is known to emit will sometimes reach harmful levels in the city. The company has produced an impact analysis in a standard way, which finds that pollutants throughout the county will remain within acceptable levels. But the analysis is at a fairly coarse spatial scale, and the residents know that the topography of the region is such that some neighborhoods in the city—smaller in scale than could be resolved by the company's analysis—sometimes experience much worse air quality than others. The city council contracts a group of scientists at a nearby university to probe in a more fine-grained way what the pollution might be like in particular neighborhoods; their overriding concern is with protecting community health.

In this example, the research is being done in a consulting mode, and the findings will inform a relatively high-stakes decision where one type of error could have serious negative consequences for the public (namely, exposure to chemicals that harm their health). This is a situation in which there is good reason to try to conduct research that aligns with the community's broader aim of protecting their health, especially if it can be done without sacrificing much epistemic quality.

Step 2. *Perform epistemic projection. The research problem that is specified must be scientifically acceptable and, ideally, will leave room for research of excellent epistemic quality.*

EPA does not require that anyone in particular participates in the process of epistemic projection. However, when research is being done to inform the decisions and actions of others, as in the air pollution example, then often it will be best for epistemic projection to be performed collaboratively by both the researchers and those actors. The researchers will have a better sense of which epistemic goals are feasible and which methodologies are available, but the actors will have a better understanding of the aims and values to which the investigation is meant to be responsive.

A helpful way to approach epistemic projection in practice is by answering a series of questions that correspond to the different elements that an epistemic research problem can have. They are articulated here for the case where there is some broader aim (reflecting contextual values) motivating the investigation, as in the air pollution example:

- 1 Which plausibly obtainable epistemic goods are particularly relevant to pursuit of the broader aim? (*goals*)
- 2 If multiple epistemic goods will be pursued, are some of them more important or relevant than others to achieving the broader aim? If so, which one(s)? (*priorities*)
- 3 Would the use of some epistemic resources (concepts, data, models, methods, etc.) rather than others be advantageous to the pursuit of the broader aim? If so, which ones? (*resources*)
- 4 Would some errors in findings, more than others, impede pursuit of the broader aim? If so, which ones? (*inductive risk*)
- 5 Would it facilitate pursuit of the broader aim to express or represent findings in a particular way? If so, which way? (*form of conclusion*)

Answers to these questions can guide the specification of one or more epistemic goals, as well as preferences and/or constraints, which jointly constitute an epistemic research problem that aligns with the broader aim.

A complication is that, since preferences come in degrees, some indication of their strength also should be provided. The best way to do this will vary with the context, but manageable approaches are readily identified. For instance, preferences might be classified as “weak,” “moderate,” or “strong,” where some qualitative description of what this means is provided. Or specific limits might be set, such as, prioritize goal A over goal B, until goal B is compromised to this extent. These sorts of approaches will not perfectly capture preferences, but EPA accepts some imperfection for the sake of tractability.

As noted earlier, when deciding whether to include a particular preference or constraint in a research problem, not only the benefits (in terms of aligning research with the broader aim) but also the costs—both epistemic and practical—should be considered. At a minimum, the research problem needs to leave room for research that respects basic constraints on epistemically adequate science. In cases where this condition is met but researchers judge that satisfying the constraints of the research problem would still compromise epistemic quality more than they are comfortable with, they should feel free to decline to perform the research. (Likewise, researchers should be able to decline to address a research problem if they judge it to have ethically objectionable motives or, more generally, if they judge that there are not actually compelling reasons to make the research responsive to contextual values in the ways reflected in a research problem.)

Example. In the air pollution example, we can imagine that the scientists and city council members together perform epistemic projection with the help of questions (1)–(5). [Table 11.1](#) shows an epistemic research problem that might be specified as a result of this process, along

Table 11.1 Example of an Epistemic Research Problem Formulated via Epistemic Projection

Research problem component	Research problem specification	Rationale
Epistemic goal(s)	Learn what the statistical distribution of pollution levels for chemicals C_1 and C_2 would be at locations around town during a typical year	Learning this is highly relevant to the aim of protecting citizens' health
Epistemic priorities	<i>Preference:</i> Prioritize accuracy of the upper quartiles of the distributions, and for neighborhoods $\{N\}$ <i>Strength:</i> Moderate; the quality of estimates of the median levels and for other neighborhoods shouldn't be greatly reduced	High-pollution days are of greatest risk to health, and neighborhoods $\{N\}$ are suspected to be at greater risk of high-pollution days; but other neighborhoods deserve quality information too, and median levels are also of interest
Inductive risk	<i>Constraint:</i> Risk overestimating pollution levels if some inductive risk must be taken on, but avoid taking on inductive risk whenever feasible, e.g., by propagating uncertainty	Overestimating pollution is better than underestimating it, given the goal of protecting health, but the council wants the results to be easily defensible if challenged by the company
Epistemic resource	<i>Constraint:</i> Use computer simulation model that can resolve pollution at neighborhood scales	That some neighborhoods may experience much more pollution than others is of particular concern
Form-of-conclusion	<i>Constraint:</i> Present findings as maps of median and 99th percentile pollution levels, and provide some indication of the associated uncertainty	Maps of middling and high levels are easier for community members to understand than statistical distributions; uncertainty information will help defend against any challenges from the company

with an indication of the rationale for each component of the problem. Notice that the research problem's specification makes no mention of the broader aim of protecting community health; it refers only to features of research qua epistemic activity. Moreover, it seems to be scientifically acceptable. In fact, carrying out research that meets the constraints and preferences in Table 11.1 might not require sacrificing any epistemic quality at all (relative to alternative routes of investigation that might be selected in pursuit of the epistemic goal).

Step 3. Try to solve the research problem

The third step occurs in—and indeed constitutes—the research stage of investigation. Researchers attempt to solve the epistemic research problem that was specified in Step 2. Solving an epistemic research problem requires, at a minimum, achieving its epistemic goals while satisfying its epistemic constraints; ceteris paribus, the more epistemic preferences are also satisfied, the better.

Example. In the air pollution study, we can imagine the university researchers making a number of choices when addressing the research problem. We can expect that they will employ a pollution model with a spatial resolution fine enough to resolve neighborhood-level variation. They might invest some time testing and improving the model's representation of a physical process that is particularly relevant for capturing the movement of pollution in the county's topography, to try to ensure that pollution levels, especially levels in neighborhoods $\{N\}$, are accurately estimated. We can expect that they will present their findings in the map format specified by the research problem. And so on.

Notice, however, that these choices by the university researchers will be justified as a means of addressing an epistemic research problem, not by appeal to any broader aims of the investigation (which, as noted earlier, in principle needn't even be known by individuals conducting research). At the same time, quite different choices might be made if, in the problem-selection stage, the broader aim of protecting community health had not been considered, and researchers were simply asked to investigate the potential pollution changes stemming from the proposed facility. In that case, the researchers might have chosen a standard lower-resolution model like the one the chemical company used, made no adjustments to the model, and presented results differently, e.g., as average pollution levels inside and outside the city limits.

11.5 Strengths and limitations of EPA

As an approach to value management, EPA has some attractive features.

First, there is EPA's potential for rapprochement. EPA recognizes that sometimes there are good reasons to make research responsive to contextual values. At the same time, EPA permits contextual values to play a justificatory role only in the problem-selection stage of investigation, and it calls for researchers to use epistemic quality as the default criterion when choosing among routes of investigation. Consequently, EPA could appeal both to individuals who reject the VFI but hold that epistemic considerations should still enjoy a qualified priority *and* to defenders of the VFI who are primarily concerned that value influence will result in shoddy science. (It will not, however, be acceptable to defenders of the VFI who hold that contextual values should *make no difference* to how research is pursued nor to challengers of the VFI who reject even a qualified epistemic priority.)

Second, even if no such rapprochement is achieved, EPA is an approach to value management that is relatively easy to implement and may be more appealing to practitioners than many alternatives. For practitioners who are eager to make some of their research responsive to contextual values but are unsure *how* to do it, EPA provides concrete guidance: They can consider a limited set of dimensions of research (goals, priorities, resources, inductive risk, form of conclusions) in the problem-selection stage of investigation, and then proceed with research in the usual way. And for practitioners who only reluctantly accept that some of their research should be responsive to contextual values, EPA may be more comfortable than approaches to value management that place less emphasis on epistemic quality.

At the same time, EPA also has some limitations.

First, EPA is a pragmatic approach to value management, not an ideal one; it sacrifices perfection for the sake of tractability. It is plausible that researchers would be less likely to miss opportunities for making research responsive to contextual values if those values were kept in view throughout the research stage and considered each time a methodological choice was made. But this would be much more onerous for practitioners. Likewise, EPA accepts that the strength of preferences regarding priorities, inductive risk, etc., identified in the process of epistemic projection typically will only approximately reflect the extent to which the anticipated users of the findings hold particular contextual values.

A second limitation of EPA is that it is incomplete in a significant way. EPA doesn't provide a general account of when it is appropriate for research to be responsive to contextual values. There seem to be some easy cases, as in the air pollution example, where research was undertaken in a consulting mode and could be significantly aligned with an ethically unobjectionable broader aim without sacrificing much, if any, epistemic quality. But matters can quickly become controversial. What if the city council only cared about pollution in the city's wealthy neighborhoods? Is it appropriate for pharmaceutical companies to implement EPA, with profitmaking as one of the

contextual values to which their research is responsive? If so, how much epistemic quality may they legitimately sacrifice for the sake of profitmaking? EPA in its current form does not directly address difficult questions like these; it does not provide a complete answer to the new demarcation problem.

That said, as noted earlier, researchers employing EPA may always decline to address a research problem that they judge to have ethically objectionable motivations or that includes constraints that can only be met by conducting research that fails to meet the researchers' own standards for epistemic quality. Moreover, when research does proceed, the transparency requirement of EPA, if faithfully implemented, will facilitate its critical evaluation.

11.6 Conclusion

EPA is an intermediate approach to value management. It recognizes that there are sometimes good reasons to make research responsive to contextual values, but it achieves this responsiveness via the careful formulation of a research problem in the problem-selection stage of investigation. EPA is thus an approach that could be acceptable to some parties on both sides of the debate over the VFI. And even if EPA fails to achieve any such rapprochement, it has other strengths. It is an approach to value management that provides concrete, tractable guidance for practitioners, and it achieves significant value responsiveness while maintaining a firm commitment to the epistemic quality of research.

References

- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Betz, Gregor. 2017. "Why the Argument from Inductive Risk Doesn't Justify Incorporating Non-Epistemic Values in Scientific Reasoning." In *Current Controversies in Values in Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge. <https://doi.org/10.4324/9781315639420-6>.
- Brown, Matthew. 2017. "Values in Science: Against Epistemic Priority." In *Current Controversies in Values in Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge Press.
- Carrier, Martin. 2022. "What Does Good Science-Based Advice to Politics Look Like?" *Journal for General Philosophy of Science* 53: 5–21. <https://doi.org/10.1007/s10838-021-09574-2>.
- DiMarco, Marina, and Kareem Khalifa. 2019. "Inquiry Tickets: Values, Pursuit, and Underdetermination." *Philosophy of Science* 86 (5): 1016–28. <https://doi.org/10.1086/705446>.
- Douglas, Heather E. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21. <https://doi.org/10.1086/674345>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5: 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Jeffrey, Richard C. 1956. "Valuation and Acceptance of Scientific Hypotheses." *Philosophy of Science* 23 (3): 237–46. <https://doi.org/10.1086/287489>.
- Lusk, Greg, and Kevin C. Elliott. 2022. "Non-Epistemic Values and Scientific Assessment: An Adequacy-for-Purpose View." *European Journal for Philosophy of Science* 12 (2): 35. <https://doi.org/10.1007/s13194-022-00458-w>.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom Is Worth Pursuing." *European Journal for Philosophy of Science* 13 (4): 48. <https://doi.org/10.1007/s13194-023-00552-7>.

- Parker, Wendy S. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36. <https://doi.org/10.1017/psa.2023.107>.
- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54: 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values in Science*, edited by Kevin. C. Elliott and Daniel Steel. Routledge. <https://doi.org/10.4324/9781315639420-4>.

DEMOCRACY, CONSENSUS, AND THE VALUE-FREE IDEAL

Tarun Menon and Jacob Stegenga

12.1 Introduction

Most philosophical accounts of the role of values in science claim that the value-free ideal (VFI) has been decisively refuted. Matthew Brown has recently claimed, for example, that “There is a general or near consensus on the view that the ideal of value-free science is untenable” (2024, 1). Several philosophers have pushed back against this status quo, including [Betz \(2013\)](#), [Hudson \(2016\)](#), and [Parker \(2024\)](#). Yet most critics of the VFI appear unmoved, claiming that these defenses of value freedom neglect the central arguments against the VFI, such as the argument from inductive risk, or that these defenses amount merely to strategies for managing the influence of values on science ([Brown 2024](#); [Elliott 2022](#)). We have articulated a novel version of the VFI which we argue is immune to existing challenges ([Menon and Stegenga 2023](#); [Stegenga and Menon 2023](#)). Our aim in this chapter is to present this version of the ideal and to offer freshly articulated arguments in its favor.

We construe the VFI as a principle intended to guide the actions of scientists, which says that scientists should do all they can to eliminate the influence of values on their reasoning. We refine this in what follows. It is commonly acknowledged that the VFI is compatible with values influencing science in many ways. [Proctor \(1991\)](#) argues that values influence science in a number of ways, such as which research topics are studied, what technologies are developed, and how research is carried out. Most conceptions of the VFI, ours included, do not rule out all of these influences of values on science. The VFI is usually understood as referring to the “internal” aspects of scientific reasoning. However, as [Elliott \(2025\)](#) points out, there are different conceptions of what qualifies as “internal” in this context. We are concerned with a very specific sense in which values may influence science: Whether they make a difference to scientific inference. We flesh out this “difference-to-inference” model in [Section 12.4](#). We argue that when values influence science in this particular way, it is always a problem calling for remedy, because this form of value-influence blocks consensus, and consensus is a constitutive aim of science ([Section 12.3](#)). However, an acceptable remedy to the problem must satisfy certain constraints, and such a remedy may not always be available. In such a circumstance, the obligation to continue searching for an adequate remedy remains, since scientists cannot rule out that the unavailability of a solution is transient.

We begin this chapter by elaborating this version of the VFI, stressing that it is a principle for guiding the action of scientists ([Section 12.2](#)). This sets the stage for our argument about consensus. We

then describe various strategies for resolving failures of consensus (Section 12.4). Constant striving toward that resolution is just what our VFI demands.

12.2 The value-free ideal as action guidance

A charge often leveled against the VFI is that it relies on a separation of science from its material context. Science is conducted by human beings subject to myriad psychological and social pressures that go beyond strict truth-seeking. It is, moreover, deeply entangled with the decision-making processes of individuals, corporations, and governments. If one takes the social reality of science seriously, the claim goes, one must admit the inevitability of value-based considerations permeating scientific inquiry. For a proponent of the VFI to avoid this value-ladenness, they must argue that the social embeddedness of scientific practice is not essential to the abstract epistemic character of science, and that we can treat science as an idealized system of theories and inferential methods which are in principle separable from the motivational psychology of real scientists and from the influence of real social and political institutions.

This criticism is well expressed by Heather Douglas when she argues that the VFI is based on “the notion that scientists are not involved in public life, that they provide no advisory functions, and that they provide no assistance for decision making,” and that “philosophers of science [defending the VFI] assumed that science was fundamentally and acceptably isolated from society” (Douglas 2009, 45–46). This conception of the VFI is a reflection of the views of the logical positivist tradition, a school of philosophy that set the agenda for a number of questions in the philosophy of science in the twentieth century, including the role of values of science. The positivists argued that when evaluating science as a knowledge-gathering enterprise, the philosopher must bracket out the sociological and psychological contingencies that get in the way of an appreciation of the epistemological core. In the words of the logical positivist Hans Reichenbach, “Epistemology... considers a logical substitute rather than real processes... [I]t will, therefore, never be a permissible objection to an epistemological construction that actual thinking does not conform to it” (Reichenbach 1938, 5–6). According to this view, the fact that real science may be sullied by its social context should not influence our philosophical conceptions of ideal science.

The view that the epistemology of science can and should be studied without consideration of its material context has few remaining adherents, and we do not wish to challenge this consensus. We agree that accounts of how science should function must be sensitive to the particular social structures and roles associated with scientific inquiry. However, this growing consensus about the need for a sociologically and psychologically informed philosophy of science has been accompanied by another growing consensus, namely, that the VFI must be abandoned, and as discussed above, these two views are often seen as deeply connected. We do not agree that paying attention to the social context of science undermines the VFI. In fact, the opposite is true: A careful consideration of the social role of science provides strong support for a version of the VFI.

Arguments for the VFI based on the role of science in society are not new. As Bright (2018) points out, the sociologist W. E. B. Du Bois was making arguments of this sort at the turn of the twentieth century. Rather than defending the VFI on the purely epistemic basis that the influence of values corrupts the reliability of scientific knowledge, Du Bois argued that for science to be a legitimate input to public policy in a democratic society, there must be public trust in the epistemic authority of science. This trust is undermined when scientific inquiry is seen to be morally or politically partisan. Du Bois thus proposes the VFI as an important constraint on public-facing scientific inquiry (as Proctor (1991) argues, this was a widespread view among sociologists of the time). Our argument for the VFI is in a similar spirit, grounding the ideal in a conception of the proper place of science in a democratic society.

The VFI, as we understand it, is not a feature of some Reichenbachian “epistemological construction,” separate from the real processes of science. Nor is it a description of an ideal state of science, one without any value influence whatsoever in the internal stages of scientific inquiry. We are agnostic about whether such a value-free state is feasible or desirable. Our version of the VFI, defended in more detail in [Menon and Stegenga \(2023\)](#), is intended as an action-guiding principle for scientists. The claim of the VFI is not that the ideal state of science is complete value freedom, it is that people involved in scientific inquiry should be trying to eliminate any difference-making influence of values on their scientific conclusions. We describe this position further in a later section, but we wish to highlight here that this is not a position about what science should be, it is a position about how scientists should behave.

Current discussions of value freedom often do not adequately distinguish between these two characterizations of the VFI. Take, for instance, the definition of the VFI in [Douglas \(2009\)](#): “the value judgments internal to science, involving the evaluation and acceptance of scientific results at the heart of the research process, are to be as free as humanly possible of all social and ethical values” (43). This definition of the VFI describes an ideal state of science, in which non-epistemic values have been eliminated from science. Presumably, this state will be the final product of some process of elimination of value-ladenness, but this process is seen as merely a means to the desirable end of value freedom. In contrast, our conception of the VFI is as an action-guiding principle governing the process of eliminating instances of value-ladenness. The ideal is understood as the process itself, not as the hypothetical end toward which the process aims. In other words, the VFI, according to us, is about the desirability of a principled commitment to value elimination in science, and not about the desirability of an end-state that excludes all values.

One might question whether there is really a significant difference between these two conceptions. If scientists should try to eliminate values wherever possible, does this not entail that a value-free state of science would be a desirable end? No, because action-guiding principles can compete with one another. A parent may be driven by the principle to increase their child’s long-term happiness wherever possible, but they may also be governed by a principle forbidding harm to other children. When faced with a situation where they could increase their child’s happiness, but only at the cost of harming another child, they might balk. Is this because they are not truly trying to maximize their child’s happiness? We think that the best way to understand their behavior would be to suppose that they are in fact motivated by this maximization principle, but they are also motivated by other principles, and when principles conflict, they need some means to adjudicate which principle should prevail. In such a circumstance, it may well be that even though the parent is motivated by a desire to maximize the child’s happiness, they would not acknowledge the state of maximal happiness as the ideal state.

As we argue in [Menon and Stegenga \(2023\)](#), similar considerations apply to the pursuit of value freedom. While we believe that scientists always have good reasons to eliminate value influence in science, they are also subject to other imperatives. There are ethical constraints on the kind of experimentation that is permissible; there are financial constraints limiting how many resources scientists can access; and there are sometimes policy-relevant constraints requiring that the output of science be action-guiding for policymakers. Even though the presence of these other constraints might mean that scientists should not eliminate particular cases of value influence, it does not mean that there are no good reasons for eliminating that influence. The VFI, on our view, remains a good reason for elimination, even if it is trumped on certain occasions by other reasons. Furthermore, even if it is trumped by other constraints in the short term, scientists should still consider the uneliminated value influence a problem and continue to attempt to eliminate that influence in the longer term, perhaps by searching for strategies that allow simultaneous fulfillment of seemingly conflicting principles.

Our view of the VFI as an action-guiding principle targeted at how scientists should behave rather than at some particular state of science allows us to avoid a powerful criticism raised by [Brown \(2024\)](#). Brown indicts a number of defenses of VFI for failing to connect with the actual practice of science and focusing instead on “philosophical abstractions with little purchase on the socially relevant parts of actual scientific practice” ([Brown 2024](#), 9). These versions of the VFI, he argues, fail to answer the crucial question of whether scientists working in the current social context should be guided in their practice by the VFI. Our view does not fail this test: Yes, the VFI as we conceive it is a principle that should guide actual scientific decision-making.

In fact, we would go further than that and turn the challenge raised by critics of the VFI around. We argue that a proper understanding of the social role of science compels us to endorse the VFI, and that it is in fact the critics of the ideal who have not adequately considered how the epistemic and social roles of science interact. In order to demonstrate this, we present an argument along the lines of that made by Du Bois, but different in its particulars.

12.3 Democracy, consensus, and value freedom

Scientific institutions exercise a uniquely powerful form of epistemic authority in modern democracies. The conclusions of scientific inquiry serve as privileged inputs in policymaking, public administration, and jurisprudence. It is an uncontroversial premise that some degree of epistemic authority for science is justified. But this premise is in apparent tension with a commitment to liberal democracy.

A widely held view in contemporary political theory is that legitimate state coercion must be justified (or justifiable) to all who are subject to it. This criterion of public justification has been described as the “moral lodestar” ([Macedo 1990](#), 78) and the “core” ([Gaus 1996](#), 4) of liberalism. There are various conceptions of public justification in the literature, but they all share a commitment to what [Vallier \(2011\)](#) calls “The Public Justification Principle”:

A coercive action C is justified if and only if each and every member of the public P has (a) conclusive reason(s) R to endorse C. (262)

This raises a worry when part of the justification for a coercive policy includes conclusions of scientific research. Fully appreciating the scientific justifications for these conclusions often requires expertise. For instance, if an energy policy mandating reductions in the emission of greenhouse gases is partially justified by the results of complex climate modeling, it is not reasonable to expect that a person outside the scientific community will understand the inferential chain justifying the assumptions of the model, or the interpretation of its output. Does this mean that the lay public does not “have” the reasons relevant to these justifications, in the sense required for the Public Justification Principle? The answer hinges on what precisely one considers the condition for an agent having a reason.

Most non-experts do not have, and cannot be expected to have, access to the reasons that scientists employ in justifying their conclusions. The reasons that every citizen can be assumed to have must be “manageable” ([Badano 2022](#)), which means that the reasoning must be graspable without overwhelming demands on citizens’ time and effort. Assessments of the proper role of science in democratic decision-making must address this manageability constraint on public justification. A number of recent contributions to the debate on the role of science in public justification, such as [Badano and Bonotti \(2020\)](#), [Reid \(2019\)](#), and [Ward and Creel \(2024\)](#), rely on the idealization that citizens can be assumed to have reasons that require understanding the evaluative standards of science, thereby not adequately addressing the constraint.

Members of the public cannot in general be expected to have good reasons for believing the conclusions of science, beyond simple trust in the reliability of expert testimony. The epistemic

complexity involved in formulating effective policy requires the public to cede some of their epistemic authority, and what it means to cede epistemic authority is to concede that some parts of the justification of policy will not be accessible to non-experts. On this issue, we agree with [Kogelmann and Stich \(2021\)](#), who argue that the proper functioning of the modern state requires expert analysis that is inaccessible to the general public.

Does this mean we must abandon public reasons liberalism, as Kogelmann and Stich suggest? We do not think so. While the arguments used to support the conclusions of scientific investigation cannot be public justifications, the use of these conclusions for policymaking can be publicly justified, and that is all that is needed to satisfy the Public Justification Principle. Even if non-experts do not have an understanding of the epistemic tools scientists use to justify their claims, they may still have good reason to suppose that these tools are developed and deployed in an institutional environment which is conducive to the production and proper use of reliable epistemic tools, and resistant to the use of unreliable epistemic tools. If non-experts have these reasons, that is enough to grant these institutions limited epistemic authority.

The fact that the authority of these institutions must be *limited* is crucial to the public justification of that authority. Authority can only be ceded in a domain where there is good reason to believe that there are experts equipped with reliable evaluative strategies that cannot be shared with non-experts due to the manageability constraint. This is the case when it comes to strictly scientific knowledge—knowledge about the behavior of empirically detectable entities and processes. Yet, this is not the case for some other forms of knowledge, most saliently moral knowledge. We are not claiming that there are no objective moral truths, nor are we claiming that there are no expert communities who aim at such truths, such as moral philosophers; we are claiming that there is no expert community that should be given the authority to determine those truths for others. Even if there is an expert community, say moral philosophers or moral psychologists, who are better positioned to arrive at moral truth than the general public, it would still be contrary to the spirit of public reasons liberalism to cede epistemic authority over those truths to that expert community, because the fundamental motivation for public reasons liberalism is respect for the moral autonomy of citizens. Democracy, on this view, requires that the authority for endorsing the moral framework undergirding political coercion must rest with the general public. Expert communities may serve as advisors to the public but cannot take over the public's role in offering and evaluating moral reasons.

We stress the difference between the roles scientific and moral knowledge play in a democracy to highlight that scientific institutions are not knowledge-gathering institutions *tout court*. Science is an institution for the generation of specifically *scientific* knowledge, which is a proper subset of all knowledge relevant to policymaking. Scientists stray from this limited remit at the peril of violating the public justification of their authority. This has been a long-standing constraint on science—Bacon claimed that the remit of science was “pure and uncorrupted natural knowledge” rather than “the ambitious and proud desire of moral knowledge” (cited in [Proctor 1991](#), 32), and the founding of seventeenth-century scientific societies across Europe was premised on the constraint, stipulated by royal authorities, that science would not study “the secrets of religion nor the affairs of the state” and that “the Natural Philosopher is to begin, where the moral ends” (cited in [Proctor 1991](#), 34). When experts directly advise the public and policymakers how to behave, as we witnessed during the Covid-19 pandemic, we observe a subsequent decline in the authority granted to experts by the public and their (now populist) representatives.

The requirement of public justification sets constraints on how scientific institutions should seek scientific knowledge. If a member of the public is to trust these institutions, she must have good reason to believe that if she had been a qualified participant in the institution, she would have good reason to endorse any policy-relevant conclusions the institution arrived at. In other words, she must have good

reason to believe that a norm of “public” justification holds *within* the scientific community. As we have argued, a non-expert is typically not equipped to assess whether a particular conclusion is publicly justified within the expert community by evaluating the justification itself. So, if she is to have good reason to believe that a norm of public justification holds within the community, it must be because the communicative structure of the community embodies that norm, such that discourse within the community is directed toward the development of justifications that can be endorsed by all members.

Jürgen Habermas has carefully developed a framework for describing a discursive community bound by norms of public justification (Habermas 1985). Habermas distinguishes communicative action from instrumental action. Instrumental action is oriented toward individual goals, and any communication is in the service of achieving those goals. Communicative action, on the other hand, is oriented toward the collective goal of consensus based on the mutual exchange of reasons. When an agent undertakes communicative action, they voluntarily take on the obligation to justify any claims they make to their peers, in a manner that aims toward reaching agreement on the validity (or lack thereof) of the claim. Thus, good faith participation in a community of communicative action implies a commitment to strive toward agreement based purely on the offering of reasons to others and the evaluation of reasons offered by others, without any individual goal extrinsic to the discourse directing the path toward consensus.

Scientific communities embody norms of public justification insofar as their discursive structure resembles that of a community of communicative action. It is the presence of this structure that gives non-experts good reason to cede epistemic authority to these communities. Habermas is explicit that communicative action aims at consensus. The reason participants in such a community engage in practices of reason-giving and receiving is to arrive at agreement. Insofar as agreement eludes them, there is more justificatory work to be done. If scientific discourse is to meet the justificatory standard of public reasons liberalism, it must have consensus as a goal. We are not here claiming that consensus is the only goal of science, but it is one of them, and moreover it is a constitutive goal. Science doesn’t aim for consensus for an instrumental reason, such as enhancing public trust. It aims for consensus because scientific knowledge (the proper aim of scientific institutions) just is the kind of knowledge arrived at through the quest for reasons-based consensus.

So, consensus is a constitutive aim of science (see Stegenga and Menon 2023 and Stegenga 2026 for independent arguments for this view). It follows that failures of consensus in science are always problems that motivate resolution. This doesn’t mean that disagreement in science is never useful; it means that even if disagreement serves as a spur to scientific development, scientists must see the resolution of that disagreement as one of their goals.

12.4 Failure of consensus

One important way in which science can fail to attain consensus, particularly relevant to understanding the influence of values on science, is when values make a difference to inference. To make this form of value-influence precise, in our earlier work we introduced two notions: “epistemic toolboxes” and “bifurcation points.” An epistemic toolbox is an agent’s set of background theories, analytic techniques, and epistemic values that serve to justify inferences from evidence to conclusions. A bifurcation point is a disagreement about what conclusion is justified by the available evidence, based on a difference in non-epistemic values. We originally formulated the notion of a bifurcation point as follows:

A bifurcation point exists when two individuals share some evidence E , hold different value sets V_1 and V_2 , and: E and V_1 justify conclusion C_1 ; E and V_2 justify conclusion C_2 ; and C_1 and C_2 are inconsistent.

(Stegenga and Menon 2023, 438)

The inconsistent conclusions that constitute a bifurcation point are a result of the different sets of values warranting differences in the epistemic toolboxes held by the two individuals. For example, suppose that the available evidence E confirms a hypothesis H to a degree x , and Sasha's value set (V_1) entails a threshold of hypothesis acceptance less than x , and so Sasha endorses H , while Masha's value set (V_2) entails a threshold of hypothesis acceptance greater than x , and so Masha does not endorse H —there would thus be a bifurcation point between Sasha and Masha.

Yet that characterization of bifurcation points was limited in one respect. Two individuals might share the same non-epistemic values but disagree about the relevant epistemic toolbox, and that disagreement about epistemic toolboxes might itself generate inconsistent conclusions. Such a disagreement would not be represented by our original formulation of bifurcation points. That limitation, though, is easily rectified by giving a more general characterization of bifurcation points, as follows:

A bifurcation point exists when two individuals share some evidence E , hold value sets V_1 and V_2 , and epistemic toolboxes T_1 and T_2 , and: E , V_1 , and T_1 justify conclusion C_1 ; E , V_2 , and T_2 justify conclusion C_2 ; and C_1 and C_2 are inconsistent.

Under this formulation, the difference in values between the two individuals can be the source of a bifurcation point, and the difference in epistemic toolboxes can also be the source of bifurcation points (and these are not, of course, mutually exclusive, since in many cases a difference in the former can manifest as a difference in the latter, as with the example of Sasha and Masha above).

This formulation of bifurcation points affords a precise way to characterize the dispute about the VFI. We start by asking: What is the appropriate response to the presence of a bifurcation point? There are, in general, two possible sources of bifurcation points, as the formulation of the notion makes clear: Bifurcation points can arise due to disagreements about epistemic toolboxes (T -disagreements), or disagreements about values (V -disagreements). We emphasize here that the values described in V are non-epistemic values—moral, social, and political values. Epistemic values—values which are held because they are considered part of the justificatory framework of science, and not for reasons extraneous to scientific justification—are part of the epistemic toolboxes.

The formulation of bifurcation points naturally suggests several categories of possible strategies that respond to the existence of a bifurcation point, corresponding to the italicized terms in the formulation:

- i Eliminate the T -disagreement, if there is one
- ii Eliminate the V -disagreement, if there is one
- iii Change E
- iv Change C

And a fifth possible kind of strategy responding to the existence of a bifurcation point is simply to

- v Ignore the bifurcation point

Given our formulation, strategies (i) through (v) are an exhaustive array of options for responding to bifurcation points. We can now return to the question we posed a moment ago: What should scientists do in the presence of a bifurcation point? We begin by rejecting two of these strategies on the grounds that they do not fit with an appropriate philosophical conception of science.

Strategy (v) is never an appropriate long-term strategy for science, because it neglects the importance of consensus in science. As we argued in the previous section, consensus is a constitutive aim of science. To employ strategy (v) amounts to resignation in the face of disagreement and thus amounts

to forswearing this constitutive aim of science. Of course, there are transient periods of science in which the existence of bifurcation points must be accepted as a matter of practical exigency. But in the longer run and at the larger community level of science, bifurcation points are always an impetus for further inquiry precisely because scientists are motivated to eliminate them.

Strategy (ii) is never an appropriate strategy for science *qua* science. To eliminate a *V*-disagreement requires the two parties to come to an agreement about the relevant values generating the disagreement. Such a resolution can, of course, be a very good thing. Suppose that Sasha's racist values constitute a *V*-disagreement with Masha, which generates a bifurcation point between them, perhaps about how to interpret a finding that purports to demonstrate race-based differences in intelligence; were Sasha to revise her pernicious values and thereby reach value-consensus with Masha, that would be good. But of course, such a resolution can be bad, as when Sasha convinces Masha to adopt Sasha's racist values. Scientists have no special expertise about matters of value and thus are not well-equipped to be attempting to modulate their own or others' values as a means to eliminate bifurcation points, as Arnon Levy argues in a compelling article (2025). Moreover, as we argued in [Section 12.3](#), the limitations set by the public reasons liberalism approach to the role of science in democracy place moral knowledge outside the proper remit of scientific institutions. For scientists to address bifurcation points by attempting to arrive at consensus about values would be to take on an epistemic role beyond what is publicly justifiable. So, while consensus about value judgments may be a good thing, sometimes, in some contexts (as [Brown 2024](#), 21 notes, channeling [Kourany 2010](#)), it is not an appropriate aim to be pursued *by science*.

Strategy (iii) is a routine aspect of science. Improving evidence, by increasing the quality or quantity of evidence, can eliminate a bifurcation point. To see this, return to the initial example of Sasha and Masha, in which the available evidence *E* confirms a hypothesis *H* to a degree *x*, and *x* is above Sasha's threshold of acceptance but below Masha's threshold of acceptance. Now suppose that they acquire more evidence which increases the confirmation of *H* to degree *y* and suppose that *y* surpasses Masha's threshold of acceptance. Both Sasha and Masha now endorse *H* and thus they have eliminated the bifurcation point that existed prior to their acquisition of the new evidence. Acquiring more and better evidence is of course epistemically beneficial, and eliminating a bifurcation point is a step toward the attainment of consensus.

Strategy (iv) is also a routine aspect of science. Scientists regularly hedge their conclusions. To see that hedging can eliminate a bifurcation point, return again to the initial example of Sasha and Masha: The evidence *E* confirms a hypothesis *H* to a degree *x*, and Sasha has a threshold of hypothesis acceptance less than *x*, and so Sasha endorses *H*, while Masha has a threshold of acceptance greater than *x*, and so Masha does not endorse *H*. Now, suppose as originally formulated *H* was a relatively precise hypothesis, say, a prediction about the change in global temperature by the year 2100 rounded to one decimal of a degree centigrade (say, "the global temperature will increase by 2.2 degrees"). To eliminate their bifurcation point, they could consider a hedged hypothesis, *H'*, which is a prediction about the change in global temperature by the year 2100 rounded to an integer degree centigrade (say, "the global temperature will increase by 2 degrees"). Suppose further that *E* confirms *H'* to a degree sufficiently greater than *x* such that Masha endorses *H'*. Thus, by hedging *H* to *H'*, Sasha and Masha have eliminated a bifurcation point. Hedging is epistemically modest, which is generally a good thing, and, just as with strategy (iii), hedging can be an effective means to attain consensus. However, strategy (iv) is not *always* appropriate, since scientists are sometimes called upon by policymakers to provide precise and informative guidance for policy, and excessively hedged conclusions may be insufficiently informative ([Frisch 2020](#); [Menon and Stegenga 2023](#); [Steele 2012](#)).

Strategy (i), while less routine than strategies (iii) and (iv), is a very important aspect of science. Eliminating *T*-disagreements requires publicly articulating the details of epistemic toolboxes,

affording criticism and refinement and occasionally the development of entirely new and powerful elements of an epistemic toolbox.

Our VFI can now be expressed as the imperative to strive to eliminate bifurcation points using strategies (i), (iii), and (iv). This formulation is slightly more general than a strict VFI per se, because some bifurcation points can be a result of *T*-disagreements which are not themselves a result of *V*-disagreements. Yet this formulation is faithful to the fundamental place of consensus in our philosophical conception of science—what matters is not merely value-based disagreements but any threat to consensus. In our earlier work we used the following slogan to express our VFI:

Scientists should act as if science should be value-free.

In that earlier work we granted that the argument from inductive risk may provide compelling grounds to require some scientific inferences to be value-laden, particularly when scientists are called on to provide guidance to policymakers and they cannot, in the short term, sufficiently deploy the above strategies. For that reason, we distinguished between end-state desirability and pursuit desirability to argue that even if the end-state of scientific reasoning should be value-laden, it could still be the case that it is good to pursue value freedom. In more recent work, not yet published, we have become more wary of the argument from inductive risk, though we need not argue the point here, since our version of the VFI stands even if one grants the merits of the argument from inductive risk. As we argued in [Section 12.2](#), the ideal is meant to be one of the guiding principles of scientific investigation, but not the only one. Conflicting principles might make the elimination of certain bifurcation points undesirable, all things considered. However, if a bifurcation point cannot be eliminated with currently available epistemic toolboxes without violating some constraint, that is not a reason for scientists to accept that the value-influence represented by that bifurcation point is permanently unavoidable. Part of the progress of science is the development of new epistemic tools, and the possibility is always open that some such tool will eventually permit the elimination of the bifurcation point while respecting the constraint. The new tool may, for instance, be a lower cost means of acquiring additional evidence that allows the elimination of bifurcation points without violating a cost constraint, or it may be a new tool for mathematical representation of the common structure of two distinct models, allowing for a hedged conclusion that is still sufficiently action-guiding. The potential of discovering such tools, combined with the orientation toward consensus constitutive of science in a democracy, mean that the elimination of any given bifurcation point should always be part of the scientific agenda.

12.5 Conclusion

Here and in previous work, we defend a new version of the VFI. Our VFI is motivated by the importance of consensus for science, is warranted by the authoritative role of science in democracies, and provides a clear action-guiding principle: Scientists should strive to eliminate the influence of values in scientific reasoning. That is because values can make a difference to inference, which thereby impedes consensus, and consensus is a constitutive aim of science.

References

- Badano, Gabriele. 2022. “Are Numbers Really as Bad as They Seem? A Political Philosophy Perspective.” In *Limits of the Numerical*, edited by Christopher Newfield, Anna Alexandrova and Stephen John. The University of Chicago Press.
- Badano, Gabriele, and Matteo Bonotti. 2020. “Rescuing Public Reason Liberalism’s Accessibility Requirement.” *Law and Philosophy* 39: 35–65. <https://doi.org/10.1007/s10982-019-09360-8>.

- Betz, Gregor. 2013. "In Defense of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Bright, Liam Kofi. 2018. "Du Bois' Democratic Defence of the Value Free Ideal." *Synthese* 195 (5): 2227–45. <https://doi.org/10.1007/s11229-017-1333-z>.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin C. 2025. "Characterizing the Value-Free Ideal: From a Dichotomy to a Multiplicity." In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Tsou, Jamie Shaw and Carla Fehr. Springer.
- Frisch, Mathias. 2020. "Uncertainties, Values, and Climate Targets." *Philosophy of Science* 87: 979–90. <https://doi.org/10.1086/710538>.
- Gaus, Gerald F. 1996. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. Oxford University Press.
- Habermas, Jürgen. 1985. *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Beacon Press.
- Hudson, Robert. 2016. "Why We Should Not Reject the Value-Free Ideal of Science." *Perspectives on Science* 24 (2): 167–91. https://doi.org/10.1162/posc_a_00199.
- Kogelmann, Brian, and Stephen G. W. Stich. 2021. "When Public Reason Falls Silent: Liberal Democratic Justification versus the Administrative State." In *Oxford Studies in Political Philosophy Volume 7*, edited by David Sobel, Peter Vallentyne and Steven Wall. Oxford University Press.
- Kourany, Janet. 2010. *Philosophy of Science after Feminism*. Oxford University Press.
- Levy, Arnon. 2025. "A (Consequence-Oriented) Critique of the Argument from Inductive Risk." *Philosophical Studies* 182: 2079–99. <https://doi.org/10.1007/s11098-025-02319-7>.
- Macedo, Stephen. 1990. *Liberal Virtues*. Clarendon Press.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom is Worth Pursuing." *European Journal of Philosophy of Science* 13 (4): 48. <https://doi.org/10.1007/s13194-023-00552-7>.
- Parker, Wendy S. 2024. "The Epistemic Projection Approach to Values in Science." *Philosophy of Science* 91 (1): 18–36. <https://doi.org/10.1017/psa.2023.107>.
- Proctor, Robert. 1991. *Value-Free Science? Purity and Power in Modern Knowledge*. Harvard University Press.
- Reichenbach, Hans. 1938. *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. The University of Chicago Press.
- Reid, Andrew. 2019. "What Facts Should Be Treated as 'Fixed' in Public Justification?" *Social Epistemology* 33 (6): 491–502. <https://doi.org/10.1080/02691728.2019.1637965>.
- Steele, Katie. 2012. "The Scientists Qua Policy Advisor Makes Value Judgments." *Philosophy of Science* 79 (5): 893–904. <http://www.jstor.org/stable/10.1086/667842>.
- Stegenga, Jacob. 2026. *Heart of Science*. The University of Chicago Press.
- Stegenga, Jacob, and Tarun Menon. 2023. "The Difference-to-Inference Model for Values in Science." *Res Philosophica* 100 (4): 423–47. <https://doi.org/10.5840/resphilosophica2023928102>.
- Vallier, Kevin. 2011. "Convergence and Consensus in Public Reason." *Public Affairs Quarterly* 25 (4): 261–79. <https://www.jstor.org/stable/23057084>.
- Ward, Zina B., and Kathleen A. Creel. 2024. "To Hedge or Not to Hedge: Scientific Claims and Public Justification." *Philosophy of Science* 91: 990–1008. <https://doi.org/10.1017/psa.2024.17>.

13

CRITICAL CONTEXTUAL EMPIRICISM

Kirstin Borgerson

13.1 Introduction

Values are present at every stage of scientific research, from the initial “light bulb” moment when an idea forms in the mind of a scientist through the identification of a testable hypothesis, selection of an appropriate methodology and mode of analysis, and dissemination and translation of results. Philosophers of science today regard it as largely uncontroversial that values are ubiquitous in science (Kincaid et al. 2007). If we take the presence of values as given within scientific inquiry, as philosophers of science suggest, a family of worries arises: Does that mean that the results of science are untrustworthy? Are we left with no means of assuring ourselves that we are closer to truth at the end of inquiry? Might the scientific facts we are presented with by scientists simply be reflections of their own prejudices or political agendas? Variations of this concern arise not only because we rely on these facts in our day-to-day lives but also because policymakers at all levels rely on scientific evidence to determine how best to structure our societies: What laws to pass, what restrictions to place on individual freedoms, what requirements to impose on professions, and so on. What can be done to alleviate these concerns?

A variety of what might be thought of as “value management strategies” have been offered by philosophers over the past half-century or so. Chapters in Parts II and III of this *Handbook* will explore others of these strategies in detail. I believe that the most promising strategies have similar features: They start from a recognition that science is a social activity, they maintain a commitment to truth or empirical adequacy as a core value within scientific communities, and they allow that different scientific domains may generate slightly different strategies for subjecting values to scrutiny. What Longino’s *Critical Contextual Empiricism* does, in particular, is offer norms that epistemic communities can adopt in order to ensure that any values operating in their domain are brought to light and critically scrutinized. The transformative critical interrogation of background assumptions and values from diverse perspectives that is recommended by CCE does not wash our scientific practices clean of values, or guarantee that values have been eradicated, but rather shines light on them and allows community members to make explicit decisions about which they will endorse and adopt and which they will reject. This allows us to benefit from the selection of values that advance our epistemic goals, and it prevents the uncritical and unexamined influence of idiosyncratic values. Let’s look at this approach in a bit more detail, and then consider some common critiques.

13.2 What is *Critical Contextual Empiricism*?

Helen Longino proposed one of the leading theories in social epistemology, Critical Contextual Empiricism (CCE), in her landmark 1990 book *Science as Social Knowledge*, and expanded her thinking on the subject in *The Fate of Knowledge* in 2002. Though there have been some minor clarifications and friendly amendments to that account over the years, it has yet to face a knock-down objection from fellow philosophers of science, or to run up against any devastating limitations in application to real-life scientific issues. Let's examine CCE by breaking down each of the components in the name.

CCE is *critical* because intersubjective critical discussion is the key mechanism by which the account identifies and scrutinizes values. And not just any critical discussion will do the trick here—the discussion needs to include a variety of perspectives and involve scrutiny of all values, including the most deeply hidden background assumptions in a community. The perspectives of people whose views are very different from those held by the majority in a community are especially highly valued in this account.

CCE is *contextual*, meaning that the standards set by any particular epistemic community are entirely its own. Physicists and clinical researchers may well adopt quite different standards because researchers in those domains have to make trade-offs between other epistemic values. For instance, simplicity may be more highly valued in physics, while generalizability may be more highly valued in the more-applied domain of clinical research. Standards are adopted and revised in light of the epistemic goals set by the community members.

CCE is *empirical* because it emphasizes the importance of empirical adequacy as a criterion of scientific knowledge. All scientific communities will share at least a commitment to empirical adequacy as one standard among their set. Longino's community norms can be thought of as socializing the process of epistemic justification. If knowledge is justified true belief, then it makes sense that she needs some version of a truth condition in order to have a full epistemic theory.

How do these components of CCE come together? Longino proposes four norms aimed at facilitating and protecting the transformative critical interrogation of values within knowledge-productive communities. The knowledge claims of a scientific community will be objective to the extent that the community upholds these norms: (1) recognized avenues for criticism (of evidence, methods, assumptions, and reasoning), (2) responsiveness to criticism, (3) shared standards that critics can invoke, and (4) tempered equality of authority (Longino 1990, 2002). Note the central role for communities in Longino's account. It is communities, not individuals, that produce knowledge.

13.3 Early critiques of CCE

Early critiques of CCE frequently included the charge that the account is insufficiently naturalistic—either that the norms seem to come from nowhere, or from intuition. Critics suggested that it would be better if the norms were derived from real scientific practices or at least if they had been subject to and survived empirical evaluation so that they might be determined to be superior to alternative norms (Solomon and Richardson 2005). A related early concern was that CCE lacks specificity: The norms are too abstract to be helpful and seem to operate at too high a level to be applied by scientists (Goldman 2002). The account was also criticized for rendering most of what was produced historically by scientists not really knowledge at all since most historical scientific communities did not enact the four norms to any significant extent (Solomon and Richardson 2005). Additionally, the account received criticism for being too permissive, because it allows even those with racist, sexist, and other problematic ideological views to intervene in scientific discussions (Hicks 2011). Finally, Longino was criticized for socializing, or democratizing, only the middle

part of the process of scientific inquiry (the method) and paying insufficient attention to the values operating at the earliest stages of inquiry (setting the research agenda) or later stages (application in practice) (Kitcher 2002).

None of these objections cut to the heart of the account or required substantial adaptation of the norms. The norms are, indeed, generated by thoughtful reflection in a manner consistent with normative thinking in other areas of philosophy. That process may not satisfy everyone, but then nothing ever does, and more thorough-going naturalists are welcome to propose and evaluate alternative accounts. Evaluating the success of the norms would require that we set a success condition or outcome, and it is hard to imagine what this could be without begging the question. Even so, Longino has always been open to these sorts of evaluations if we can figure out how to do them. Specification of the norms in light of evolving scientific practices makes sense, but it ought to be context dependent since not every community faces the same pressures or challenges. For example, clinical researchers in the early 2000s needed to find a way to deal with a growing tendency of researchers in that domain to simply refuse to publish their results (a phenomenon with many names, but usually called *publication bias*). Researchers were motivated to bury their research results for a variety of reasons, including industry sponsorship and financial gain. A minor modification to the norm of recognized avenues for criticism, like requiring that scientists submit their research for critical examination in the recognized venues in a timely manner and in an accessible format, provided a helpful specification of the norm that was responsive to the contextual features of that situation (Borgerson 2011). In this way, specification of the norms can be attentive to developments in a field. And if the specification is in keeping with the spirit of CCE—it is aiming to protect transformative critical discussion against a common threat—it may well be worth adjusting or adapting the norms at the general level.

Further, the norms are abstract because this is fundamentally a contextual account and Longino is trying to hold space for the incredible variety of practices in different domains of science. Expecting more specificity from a contextualist misses the point of contextualism. I explore this response further when we discuss a contemporary criticism below. Turning to the historical concern, scholars can simply bite the bullet: Historical scientific practices may well have been less knowledge-generating, and more deeply value-laden, than appreciated. Perhaps if science of the past was more open to different perspectives, we would have fewer racist, sexist, ableist, ageist, homophobic, etc. scientific views reflected in textbooks from the past century (Comen 2024). This brings us quite naturally to Hicks' concern: Longino's account is deliberately inclusive of diverse perspectives, but this inclusivity may extend even to people that Longino (as a feminist scholar) would deeply disagree with, such as those with racist, sexist, and other oppressive ideologies. This is indeed possible under her account. In response to this concern, scholars have pointed out that the four norms, working in conjunction, might nonetheless protect scientists from wasting too much time in exchanges over such views. The responsiveness norm, for instance, requires that all parties, including those representing outsider views, be responsive to feedback. Insofar as ideologically committed scholars are unable or unwilling to revise their own views, they will run afoul of this norm and can be set to the side (Rolin 2017). If the sexist or racist scientists employ more subtle approaches, such as assuming negative stereotypes about oppressed groups in the design of research questions, CCE is still our best option for subjecting such views to critical scrutiny and rejecting them because they are exposed rather than hidden. (The value-free ideal has certainly failed at this task, given the many and ongoing case studies of sexist science provided by feminist philosophers of science.) Kitcher's point about the limitations of the account is perhaps the most interesting in the set but again seems mainly to be a criticism of Longino for engaging in one project (finding a way to deal with the values that infiltrate the practices of sciences throughout) rather than another (correcting unfairness in the distribution of scientific resources). There is no reason why CCE couldn't be adopted alongside strategies to democratize or otherwise improve the fairness of early and later stages of inquiry.

A more contemporary concern relates to the tactics used by some researchers to manufacture uncertainty or doubt (Fernández Pinto 2014, 2015; Jukola 2015; Michaels 2008). Tactics of subjecting near-consensus views in science, such as views about the harms of tobacco, to “further research” are often used to preserve the market for such goods as long as possible. They are not aimed at improving the objectivity of our knowledge claims through critical interrogation, but they might be presented as if this is what they are doing and appear therefore to be endorsed by CCE. As I argued in my 2011 paper, manufactured doubt may well impede our ability to assign the honorific of knowledge to a particular claim (we may need to await assessment of the new evidence—however manufactured), but this doesn’t stop policymakers from moving ahead with policies based on what we have very good reason to believe (Borgerson 2011). Tactics of manufactured uncertainty need not be an impediment to good policy, even if the best response epistemically is to permit the delay.

Finally, the most recent criticisms tend to echo either the earliest criticisms noted above or present a development in some domain of science and question whether the norms are still appropriate to provide helpful guidance given this development (cf. Koskinen 2017). There is a common mistake made in these papers, which is to blame CCE rather than the scientific communities in question for the failings or limits of that new scientific development. If the worry is the emergence of groupthink in interdisciplinary scientific communities, this isn’t a problem for CCE (as is claimed), *it is a problem for those communities in which groupthink is emerging*. CCE is doing exactly the normative work we would hope it would do in drawing our attention to the ways in which practices like groupthink make the critical interrogation of background assumptions more challenging, and thus objectivity harder to achieve. This is worth acknowledging alongside our appreciation for some of the benefits of interdisciplinary teamwork in the sciences. It is the communities that ought to be the target of criticism, not CCE. With this overview and a few cautionary tales in mind, let’s turn to one example of recent critical engagement with CCE and see whether it fares better than some of these earlier efforts. I will consider and respond to this objection in detail, in order to demonstrate how a contemporary exchange over CCE might unfold.

13.4 Recent critique of CCE

In a paper entitled “What Counts as Relevant Criticism?,” Teemu Lari claims to have identified a neglected problem with CCE, namely that it contains two contradictory principles (2024). Lari suggests that failure to clarify the principles in question renders CCE unable to make sense of an important type of criticism in scientific exchanges: Goal-directed criticism. He proposes revisions to CCE aimed at clarifying the principles so as to avoid conflict.

The first principle, which Lari labels the “openness of goals to criticism” (OGC) principle indicates that “the cognitive goals of epistemic communities may be criticized, and this criticism must be responded to by defending or revising the criticized goals” (Lari 2024, 90). Lari provides extensive textual support for this principle within Longino’s work, and it is certainly consistent with the core tenets of CCE. The aim of critical interrogation is to expose the values underpinning knowledge-productive activities. The values that have shaped the *goals* of inquiry in a domain should, if anything, receive the highest level of scrutiny given their outsized effects on all that follows. What it is that we seek to know in many domains of science is currently shaped by competing values such as alleviating suffering, contributing to the survival of humanity, and financial benefit for investigators. We can’t hope to make choices among the various available goals and values until our current value-laden goal is first recognized and then critically examined.

The second principle, called the “goal-relativity of response-requiring criticism” (GRC) principle, indicates that “criticism creates a duty of response if and only if it is relevant to the goals of the community that is the target of the criticism” (Lari 2024, 90). So, for example, criticism of a philosophical

project on the ethics of euthanasia for failing to answer empirical questions about the demographics of people who access euthanasia in a particular jurisdiction does not generate a duty of response by the philosophers because it fails to appreciate the goals of normative ethical inquiry. Other examples might include criticisms of clinical researchers for failing to address economic or political questions, or criticisms of physicists for failing to answer questions about ecosystems. These sorts of misplaced criticism may be deemed irrelevant to the goals of inquiry in the target domain and ignored, without fear of violating the norms of CCE. This is a pretty common approach within academia, where questions deemed inappropriate to a particular discipline are rejected on the grounds that they miss the point of the investigation. The GRC is derivable, according to Lari, from the commitment within CCE to shared standards as a demarcation criterion for communities. Shared standards are, in turn, selected by a community on the basis of its epistemic goals or aims.

Lari argues that CCE is committed to both principles, and that the principles are in tension with one another. If, following the OGC, a researcher criticizes the goals of a particular knowledge-productive community with an expectation that the targeted researchers will respond in some way to the criticism, they are likely to be disappointed since the GRC seems to license ignoring such criticism. CCE, then, seems to both encourage and discourage criticism of the goals of inquiry in a particular community.

Lari's solution adapts the popular philosophical strategy of resolving tension by identifying two different levels or orders of discussion: First- and second-order discussion. First-order discussion, or scientific discussion proper, concerns "the appropriateness of methods, assumptions, and other commitments *in light of a given scientific project or cognitive goal*" (Lari 2024, 93 emphasis added). In this type of discussion, the goals or aims of inquiry are taken as given and criticism proceeds without further examination of those goals. Lari suggests that the GRC principle operates at this level. In other words, researchers can proceed without having to respond to criticism about anything other than their methods, analysis, and results.

Second-order, or meta-level discussion concerns "the desirability of the ends in themselves" (Lari 2024, 93). One might, for instance, critically interrogate the goals of inquiry in light of evolving social developments or needs. The OGC principle operates at this level. For example, a subset of clinical researchers (for example, feminist clinical researchers) may argue that broader and more representative populations of patients ought to be included in clinical research in order to improve the generalizability of results, and this might require a shift in the goals of clinical research.

Let's look closer at this example, because I think it helps us to see the limits of Lari's proposal. For the latter half of the twentieth century, clinical research on new treatments mainly investigated the *efficacy* of treatments (Cartwright 2011). It did so by testing treatments on carefully selected human subjects: Otherwise healthy, relatively young, typically male, populations (Comen 2024). Researchers aimed to answer the question "does this treatment work?" by testing them in what was regarded as the standard human body at the time, uncomplicated by comorbidities, hormonal fluctuations, pregnancy, complications of aging, etc. Of course, what was really being asked and answered in such studies was the question "does this treatment work in this population under these conditions in this context?" but results were treated as if they offered a universal human answer to the question about whether the treatment worked full stop.

With the benefits of a less misogynistic (not to mention less racist, ableist, and so on) perspective today, we see the obvious shortcomings of this approach (Criado-Perez 2019). We can also track the extensive harms done to women and other underrepresented groups by the uniform prescription of treatments without attention to the differences posed by physiology, social location, or age (Duseberry 2018). As a result of increasing awareness of the harms resulting from the older approach to clinical studies, some clinical researchers in the past decade or so have attempted to shift the field away from efficacy studies to *effectiveness* studies (Loudon et al. 2015). This shift represents a

change in the goals of inquiry within this domain, reflecting greater priority given to broadly generalizable knowledge achieved through more representative inclusion of patients in research, as well as more representative patient settings, contexts, and resources. We see this in the shift over the past decade toward pragmatic randomized controlled trials (RCTs), which insist that all elements of research align with practice in ways that place high value on the generalizability of results to the full spectrum of “real life” patients.

When feminists and other scholars pointed out the problems with generalizing from a narrowly defined “standard human” to all people’s bodies, clinical researchers were forced to consider what it was they were aiming at when they were conducting research, and whether their methods were a good fit with that goal. What they noticed is that their methods didn’t seem to be achieving their goal, insofar as the goal of clinical research is improvement of human health for all human beings regardless of sex, race, social location, etc. They had been over-valuing a certain kind of simplicity in their research, and under-valuing the generalizability of results to the population in question. This led (at least some of) them to rethink their goals and the methods required to achieve them. The shift toward more pragmatic research design involved both a re-evaluation of the goals of clinical research and a rethinking of the methods need to achieve those goals. One thing is hopefully clear from this example: *Criticism of the goals of research and the methods used to achieve those goals are inextricably intertwined*. One cannot criticize a failure to produce generalizable results without considering what the methods in use were capable of doing or might be repurposed to do. I believe this deep connection between methods and goals renders the first- and second-order distinction unhelpful and is likely the reason why scholars have not described the CCE commitments as inconsistent previously. If we try to impose a first- and second-order distinction here between different types of criticism, it won’t help actual communities to figure out when they should double down on their goals and when they should subject them to scrutiny. It is better to leave all elements of inquiry open to scrutiny.

Scholars interested in CCE may find it helpful to note common features of criticisms arising in the literature, since these patterns do seem to emerge and re-emerge with time. Lari’s argument, for instance, bears resemblance to one provided by Alan Goldman in his review of Longino’s 2002 book *The Fate of Knowledge*. Goldman (2002) suggests that a certain sort of dogmatism (akin to religious dogmatism) is necessary to the function of science:

Isn’t science also “dogmatic” in insisting on scientific or statistical methods? Researchers aren’t invited to challenge those methods when they submit their research papers. Indeed, general questions of scientific standards are usually relegated to philosophy journals rather than published in scientific journals. (2148)

Goldman seems to be tempted by something like Lari’s first- and second-order distinction in this quotation, whereby scientists ought to be able to go about the business of science without being subject to criticism about their foundational assumptions or goals into question at every moment.

Goldman argues that any attempt to interpret the CCE norms in such a way as to exclude religious dogmatists from being recognized as members of knowledge-productive communities will also exclude (presumably benevolent) scientific dogmatists. In response to this argument, in a 2011 paper I began by pushing back on this understanding of dogmatism: “Standards are dogmatic when people fail to have reasons for them and fail to give reasons to others who inquire about them. Scientific standards are not, or at least should not be, dogmatic in this sense” (Borgerson 2011, 443). Holding standards constant during the course of research is a perfectly acceptable thing to do. It doesn’t mean you refuse to discuss those standards when they are questioned.

In my response I also noted the familiarity of this objection: “Contextualists often face a version of this charge from foundationalists. The reply is straightforward, though not always entirely satisfying

to foundationalists—a community can and will provisionally accept certain methodological rules or background assumptions in the search for knowledge” (Borgerson 2011, 443). Foundationalists seem perpetually surprised that contextualists are satisfied with the shifting sands provided by provisional standards. Surely a firm foundation is preferred! But contextualists see the alternative as problematic for precisely the reasons Goldman provides: Foundations that are beyond question by community members, that is, are held as dogmatic by the members of the community aiming to produce knowledge, are a threat to our assurance that the community is genuinely knowledge-productive. This includes the goals and aims of a community and certainly includes the many decisions required in the design and selection of appropriate methods in research. Again, this provides us with strong reasons to avoid any artificial separation of first- and second-order discussions, or any relegation of particular types of community members to different domains or different journals. Methodological questions invoke both practical issues (which patient population will be targeted? Is the legacy of distrust among that population a barrier to statistical significance, thus requiring a different design or statistical approach?) and questions about the goals of inquiry (are we trying to answer the question “does this work?” or the question “does this work for this population, when delivered in this way, in this context, with these supports, at this dose...?”). There is no easy way to separate these issues, and nothing to be gained by doing so.

13.5 Expansion and application of CCE

In this section I will briefly outline three areas of possible expansion and application of CCE, though there are undoubtedly many more. In her 2023 paper “On the Harms of Agnotological Practices and How to Address Them,” Immaculada de Melo-Martín engages with the growing philosophical literature on ignorance. When we think of our most reliable knowledge-productive communities, scientific communities come immediately to mind; this is no doubt why they are the focus of so much of the work in social epistemology. But in addition to producing knowledge, such communities are also capable of creating misinformation, deepening our ignorance, and sowing seeds of doubt. de Melo-Martín argues that focusing on the common strategy of identifying the bad motives behind various ignorance-producing practices is insufficient to address these problems. In fact, an overzealous effort to seek out sources of misinformation may well miss the deepest issues at play here, which often have more to do with values and trust. The way forward, according to de Melo-Martín, focuses on cultivating warranted trust: “tackling institutional and social factors that undermine [warranted trust] and promoting mechanisms that contribute to ensuring the trustworthiness of scientists” (2023, 10). CCE doesn’t say much directly about trust and trustworthiness, or about how to combat the scientific production of ignorance. And if ignorance is not merely what arises from a failure to enact the CCE norms but is the result of more complex mechanisms, it isn’t clear if or how CCE can help us in this domain. Given the prominence of these topics in philosophy of science today, we might hope more could be said.

I have a few initial thoughts on this topic, though again I see this as an area for possible future research. Recall that the original motivation for developing CCE was the problem of unmanaged values in science. Identifying community-level mechanisms for ensuring critical discussion of values addresses this problem, and in doing so allows communities to be assured of the objectivity of their beliefs. Why is objectivity valuable? When we are as assured as we can be that idiosyncratic values have been identified and scrutinized (and sometimes rejected) through a process of critical interrogation, we are in a better position to trust the knowledge produced. Claims to scientific objectivity will otherwise mask values that necessarily persist in our claims to know various purported facts. There are, it seems, clear connections that can be made between the sort of objectivity pursued by CCE and trustworthiness. Suggestions for tackling the problem of scientifically produced ignorance are likely

to be in line with CCE norms. Whether the agnotological practice involves casting doubt on otherwise well-established scientific claims (as in the case of tobacco and fossil fuels research) or creating false consensus (as in the case of the sugar industry creating consensus about the harms of fat and cholesterol in the 1960s and 1970s), the remedy ought to involve critical interrogation of claims being made (de Melo-Martín 2023). There is plenty of room for scholars interested in CCE to contribute to discussions about how we might go about improving the trustworthiness of scientific institutions.

A second domain of potential interest to scholars of CCE is the heated debate currently underway over the use of artificial intelligence (AI) technologies in scientific research. Generative AI systems such as ChatGPT notoriously place a “black box” between input and output that is permanently sealed because randomization at each step within the process makes it one that cannot be replicated or even reverse engineered: “AI black boxes refer to AI systems with internal workings that are invisible to the user. You can feed them input and get output, but you cannot examine the system’s code or the logic that produced the output” (Bagchi 2023). If black-box types of AI are used as part of a scientific knowledge production, this necessarily leaves everything in the black box beyond the reach of critical interrogation. It is not hard to see the ways in which this might be problematic according to CCE. Things get even worse when we consider that generative AI systems are trawling the internet, and the internet is awash with racism, sexism, ableism, etc., which, without human filters, find their way into the outputs of such systems. CCE was proposed as a solution to the problem of values in science and the level of concern about how big an impact such values might have has never been higher due to advances in AI technology and increasing use of such technologies by scientists. CCE might be used to point to the ways in which AI-generated knowledge makes critical interrogation of background assumptions impossible. This can, at the very least, be added to the enormous heap of worries we already have about these technologies, and perhaps scholars of CCE could suggest some provisional interventions or corrections, at least in the domain of science.

A third, and perhaps the most fruitful, area for expansion of CCE is in reflections on the developments of pandemic science and “emergency” science more broadly. Scientific research conducted during the COVID-19 pandemic had a number of odd features, including different roles for laypeople in advocating for certain methods or treatments (as seen in debates over challenge trials for vaccines), rapid dissemination of anecdotal evidence through social media (sharing “what works” via Twitter/X, for instance), expedited research timelines and sharing of research results, issues around the adjudication of relevant expertise for policy decisions, and plenty of discussion about uncertainty (Lipworth et al. 2023). All of these issues (and more) would benefit from thoughtful evaluation by proponents of CCE, who might be interested in whether these changes to scientific practice bring it closer in line, or further from alignment with, the CCE norms (Canali and Lohse 2024).

13.6 Conclusion

This chapter began with a worry about the ways in which the values pervasive in science might undermine its capacity to produce trustworthy, reliable, objective knowledge. The suggestion made by Longino and other scholars of CCE is that transparency and critical interrogation of values is our best value-management strategy and is best enacted and assessed at the community level. There are other management strategies we might consider, but I hope the arguments of this chapter have convinced you that CCE is at least a strong contender among this set. CCE is a tremendously resilient theory, and its contextualism, when properly appreciated, is a strength rather than a shortcoming. Whichever value management strategy we end up with, though, one thing is clear to everyone: A return to the value-free ideal would be a mistake. As Elizabeth Anderson points out, “Insistence on the value-neutrality of science is self-deceptive and unrealistic” (Anderson 2012). Pursuing that strategy is akin to burying our head in the sand and pretending values aren’t present when they so obviously are. It

prevents us from seeing the values that necessarily remain in science and, worse, it leads us to trust science that does, in fact, reflect personal prejudices, industry influence, or political agendas. This is far more worrisome than the more transparent critical approach recommended by CCE.

References

- Anderson, Elizabeth. 2012. "Feminist Epistemology and Philosophy of Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=feminism-epistemology>.
- Bagchi, Saurabh. 2023, May 22. "What Is a Black Box? A Computer Scientist Explains What It Means When the Inner Workings of AIs Are Hidden." *The Conversation*. <https://theconversation.com/what-is-a-black-box-a-computer-scientist-explains-what-it-means-when-the-inner-workings-of-ais-are-hidden-203888>.
- Borgerson, Kirstin. 2011. "Amending and Defending Critical Contextual Empiricism." *European Journal for Philosophy of Science* 1 (3), 435–49.
- Canali, Stefano, and Simon Lohse. 2024. "How to Move Beyond Epistemic Battles: Pluralism and Contextualism at the Science-Society Interface." *Humanities and Social Sciences Communications* 11: 98. <https://doi.org/10.1057/s41599-023-02561-6>.
- Cartwright, Nancy. 2011. "A Philosopher's View of the Long Road from RCTs to Effectiveness." *The Lancet* 377 (9775): 1400–1. [https://doi.org/10.1016/S0140-6736\(11\)60563-1](https://doi.org/10.1016/S0140-6736(11)60563-1).
- Comen, Elizabeth. 2024. *All in Her Head: The Truth and Lies Early Medicine Taught Us about Women's Bodies and Why It Matters Today*. HarperCollins.
- Criado-Perez, Caroline. 2019. *Invisible Women: Data Bias in a World Designed for Men*. Harry N. Abrams.
- de Melo-Martin, Immaculada. 2023. "On the Harms of Agnotological Practices and How to Address Them." *International Studies in the Philosophy of Science* 36 (3): 211–28. <https://doi.org/10.1080/02698595.2023.2178838>.
- Dusenbery, Maya. 2018. *Doing Harm: The Truth about How Bad Medicine and Lazy Science Leave Women Dismissed, Misdiagnosed, and Sick*. HarperCollins.
- Fernández Pinto, Manuela. 2014. "Philosophy of Science for Globalized Privatization: Uncovering Some Limitations of Critical Contextual Empiricism." *Studies in History and Philosophy of Science, Part A* 47: 10–7. <https://doi.org/10.1016/j.shpsa.2014.03.006>.
- Fernández Pinto, Manuela. 2015. "Commercialization and the Limits of Well-Ordered Science." *Perspectives on Science* 23 (2): 173–91. https://doi.org/10.1162/POSC_a_00166.
- Goldman, Alvin I. 2002. "'Knowledge and Social Norms.' Review of *The Fate of Knowledge* by Helen Longino." *Science* 296: 2148–9.
- Hicks, Daniel. 2011. "Is Longino's Conception of Objectivity Feminist?" *Hypatia* 26 (2): 333–51. <https://doi.org/10.1111/j.1527-2001.2010.01160.x>.
- Jukola, Saana. 2015. "Longino's Theory of Objectivity and Commercialized Research." In *Empirical Philosophy of Science: Introducing Qualitative Methods into Philosophy of Science*, edited by Hanne Andersen, Nancy J. Nersessian and Susann Wagenknecht. Springer.
- Kincaid, Harold, John Alison, and Dupré Wylie. 2007. *Value-Free Science? Ideals and Illusions*. Oxford University Press.
- Kitcher, Philip. 2002. "The Third Way: Reflections on Helen Longino's *The Fate of Knowledge*." *Philosophy of Science* 69: 549–59. <https://doi.org/10.1086/344617>.
- Koskinen, Inkeri. 2017. "Where Is the Epistemic Community? On Democratization of Science and Social Accounts of Objectivity." *Synthese* 194: 4671–86. <https://doi.org/10.1007/s11229-016-1173-2>.
- Lari, Teemu. 2024. "What Counts as Relevant Criticism? Longino's Critical Contextual Empiricism and the Feminist Critique of Mainstream Economics." *Studies in History and Philosophy of Science* 104: 88–97. <https://doi.org/10.1016/j.shpsa.2024.02.005>.
- Lipworth, Wendy, Ian Kerridge, Cameron Stewart, Diego Silva, and Ross Upshur. 2023. "The Fragility of Scientific Rigour and Integrity in 'Sped up Science': Research Misconduct, Bias, and Hype and in the COVID-19 Pandemic." *Bioethical Inquiry* 20: 607–16. <https://doi.org/10.1007/s11673-023-10289-w>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Longino, Helen. 2002. *The Fate of Knowledge*. Princeton University Press.
- Loudon, Kirsty, Shaun Treweek, Frank Sullivan, et al. 2015. "The PRECIS-2 Tool: Designing Trials That Are Fit for Purpose." *BMJ* 350: h2147. <https://doi.org/10.1136/bmj.h2147>.

- Michaels, David. 2008. *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford University Press.
- Rolin, Kristina. 2017. "Can Social Diversity Be Best Incorporated into Science by Adopting the Social Value Management Ideal?" In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Solomon, Miriam, and Alan Richardson. 2005. "'A Critical Context for Longino's Critical Contextual Empiricism' Review of *The Fate of Knowledge* by Helen Longino." *Studies in History and Philosophy of Science* 36: 211–22. <https://doi.org/10.1016/j.shpsa.2004.12.014>.

VALUES IN GLOBAL SCIENCE AND THE RELEVANCE OF GEOGRAPHIC DIVERSITY

Juliana Gutiérrez Valderrama

14.1 Introduction

Most philosophical frameworks have posed social and epistemic diversity as a necessary feature for addressing the value-laden nature of science. A socially diverse community (with different social identities, social locations, or social values) will most likely be epistemically diverse, i.e., will have different background assumptions, worldviews, concepts, research heuristics, thinking styles, etc. (Page 2017). Accordingly, it can give room to new (and more adequate) questions, theories, concepts, hypotheses, and ontologies (Harding 2015), as well as more criticism, accountability, and a more efficient division of cognitive labor (Longino 2002; Muldoon 2013; Pesonen 2022). These elements are crucial for identifying, managing, and justifying the background assumptions and values that shape all stages of the production of scientific knowledge. In addition, a socially diverse community can lead to a more socially responsible and epistemically just science (Intemann 2009; Kourany 2010).

While much of the literature on diversity has been insightful, it has primarily focused on abstract, national, or local/institutional contexts (Koskinen and Rolin 2021) and mainly on scientific communities located in resource-rich environments in the so-called Global North (Dasgupta 2021). Philosophical discussions have revolved around matters such as how much and what kind of diversity and dissent is desirable or epistemically beneficial (Biddle and Leuschner 2015; De Melo-Martin and Intemann 2013), what institutional and structural conditions we should have to maximize the benefits of diversity (Fehr and Jones 2022; Steel and Paier 2022), how to avoid free-riding the benefits of diversity and discharging all the costs and burdens on members from minority and oppressed groups (Berenstain 2016; Fehr 2011; Steel and Paier 2022), how to increase diversity through “top-down” measures (Koskinen 2022), and knowledge integration and epistemic pluralism (Ludwig, El-Hani, and Gatti 2024; Wylie 2015).

Unfortunately, the normative tools at hand remain insufficient when we move on to explore matters of diversity beyond these boundaries and onto a global scale. Thus, various questions continue to be neglected: How can we address the value-laden nature of global science, given the existing inequalities among scientific communities in different geographical locations? If indeed diversity is a promising avenue for identifying, managing, or deciding on the values that shape global scientific knowledge and encouraging a socially and epistemically just science, what sort of diversity should we foster, and how? In particular, the philosophical debate thus far has ignored an aspect that I claim to be crucial for addressing the challenges of value-laden global science: The *geographic* dimension

of diversity. By *geographic diversity*, I understand “diversity in the locations where people [...] live and work” (Angus et al. 2021, 255). These locations, furthermore, are (i) embedded in a network of power relations or *world-system* (Wallerstein 2004) and (ii) linked to epistemic, cultural, social, racial, political, and economic imaginaries (mostly shaped by histories of colonialism) (Quijano 2000). A geographically diverse scientific community would be one where its members contribute to the production of scientific knowledge from different geographical locations across the world. This dimension, often overlooked, plays a significant role in shaping the values and perspectives that influence global science.

The chapter is divided in three sections. First, I examine how geographic diversity has been neglected or inadequately addressed in philosophical frameworks of value-laden science. Second, I provide social and epistemic rationales for considering this dimension of diversity. I intend to show how different forms of *epistemic oppression* hinder the participation of scientific communities located in low- and middle-income countries (LMICs) or the so-called Global South.¹ This affects (i) the epistemic agency of scientific communities in these locations and (ii) the values that shape the production, content, and uses of scientific knowledge globally. Finally, I conclude by pointing out what a philosophical framework must keep in mind when considering geographic diversity and conceiving measures for increasing it in a judicious way.

14.2 Geographic diversity in the current literature on science and values

There are two approaches to social diversity in current frameworks of value-laden science: *Democratic* ideals and *standpoint theory* ideals. I argue that both types of frameworks neglect the relevance of geography in various ways and, hence, are inadequate for facing and managing the values and interests that shape global science.

On the one hand, according to democratic ideals, scientific communities should (i) include researchers with different values and social identities and (ii) provide equal conditions for criticism and deliberation. This would allow us to produce a more empirically adequate, accountable, and socially relevant science (e.g., Kitcher 2001; Longino 2002). However, the emphasis of these frameworks remains on the social identity (i.e., race, gender, class, and ethnicity) and background assumptions of who belongs to the scientific community but not on where that scientific community is located. Accordingly, the world of science and knowledge circulation, enforced with our imaginaries of globalization and cosmopolitanism, becomes a “leveled playing field” or, better said, a world of “placeless flows” (Agnew 2007, 141). Using Agnew’s words, this image “presumes total ease of movement, timelessness, no directional bias, and an Archimedean view over the whole” (Agnew 2007, 140).

When we push democratic frameworks to address the situation of science and knowledge production in resource-poor conditions and subaltern regions of the world, two difficulties arise. First, this “placeless” image of science, in reality, describes very specific contexts of global science, i.e., encounters among equal scientific communities in terms of trust, authority, and resources and with a long history of intellectual exchange (common among high-income countries or the “Global North”) (Dasgupta 2021). Second, the calls for diversity and inclusion of epistemic communities outside of this “placeless” scientific exchange are generally understood in terms of engaging with and giving uptake to “local,” “indigenous,” “traditional,” or “ethno-knowledge” from “non-Western” peoples in LMICs (most of them former European colonies) (e.g., Lacey 2021). This leaves out of the philosophical discussion the scientific communities in subaltern places and in far from ideal conditions.²

On the other hand, standpoint theory ideals stress the need for the uptake of marginalized perspectives and values, especially those that maximally differ from hegemonic epistemic frameworks (e.g., Harding 2015). According to this approach, objectivity is enhanced by “starting off thought” from the perspective of those who have been oppressed since these standpoints prompt a “strong reflexivity”

(Harding 1993). However, these proposals tend to identify (“Western”) academic science as the hegemonic epistemic practice proper of privileged communities in the “Global North.” Naturally, they translate the imperative of including and developing the marginalized perspectives from subaltern regions mainly as giving uptake to “non-Western” or “Southern” epistemic practices—again, those knowledges usually labeled “local,” “indigenous,” “traditional,” etc. In Harding’s (2018) words, “[it is] precisely some of the elements of Southern cultures most incompatible with modern sciences and their philosophies that would be valued” (53).

Therefore, we find ourselves with the following outcome: Science (as traditionally understood) is mainly *situated* in those privileged locations conceptualized as “the West” or “the North,” and “Northern” scientific communities have the epistemic and social duty to engage with and uptake the “Southern” sciences from “the Rest.” Unfortunately, this excludes the situation of scientific communities from these regions, as if (“Western”) science was *alien* to these places, or as if their conditions for contributing to and accessing scientific knowledge were not problematic or relevant to science and the philosophy of science.

In sum, both democratic and standpoint theory ideals encourage us to consider geographic diversity and giving uptake to communities in subaltern regions as a matter of transdisciplinarity or integration (sometimes radically) different knowledge systems produced by culturally different social groups (e.g., Albuquerque, Ludwig, and Feitosa 2021)—where the so-called Global South is the main source of such radical difference. This may be one of the reasons, as Koskinen and Rolin (2021) correctly note, that structural epistemic injustices within global science remain unaddressed in the philosophy of science in comparison to the epistemic injustices towards other non-academic ways of knowing (Grosfoguel 2013). Besides potentially exoticizing the epistemic communities and practices of LMICs that fall under the term “Global South,” these approaches also neglect the structural inequalities that scientific communities in subaltern regions endure.

Numerous empirical studies have described the hierarchical division of global cognitive labor and the institutional and material barriers that scientific communities in LMICs face when contributing and having access to scientific knowledge. The philosophy of science has yet to pay attention to the nature of these inequalities, their epistemic and social implications, and possible measures or tools for counteracting them.³ I wish to move forward in that direction and examine these inequalities as cases of *epistemic oppression*. In what follows, first, I present the concept of “epistemic oppression” and its effects on epistemic agents and communities; next, I use this framework to examine the situation of global science.

14.3 Epistemic oppression

Epistemic oppression is a structural harm that deprives the members of a particular group of the means and conditions necessary to develop and fully exercise their epistemic agency (Dotson 2014; Tanesini 2022; Wolf 2020). These epistemic wrongs cannot be traced to individual discriminatory actions caused by identity prejudices.⁴ On the contrary, they are the product of the unintended consequences of numerous individual actions of subjects who participate in the production of knowledge within a social structure. Even if individuals are morally good and behave fairly in their epistemic exchanges, they could be causally contributing to epistemic oppression.

Following Young’s (1990) taxonomy of oppression, epistemic oppression can also adopt different faces (Tanesini 2022; Wolf 2020). *Epistemic exploitation* occurs when one party benefits from another’s epistemic labor and resources, hindering the latter’s agency and interests (Berenstain 2016; Fehr and Jones 2022). *Epistemic marginalization* denotes those situations where agents are excluded from the spaces of knowledge production, affecting their contribution and use of epistemic resources (Dotson 2014). *Epistemic powerlessness* involves cases where agents cannot determine or choose

their tasks in the division of cognitive labor and their contribution to the production of knowledge. *Epistemic violence* refers to practices of silencing, such as attacks, harassment, intimidation, humiliation, or mockery (Dotson 2011; Fernández Pinto and Leuschner 2026). Finally, *cultural imperialism* (or *epistemic subordination* (Tanesini 2022)) occurs when the epistemic resources and practices of dominant groups are set as the norm or the standard of “good” knowledge. Consequently, oppressed groups’ epistemic resources and practices are identified as deviant, inferior, or derivative.

These structural injustices have epistemically detrimental implications for knowers and epistemic communities. They can foster epistemic vices of inferiority and superiority depending on one’s position in structures of oppression (Tanesini 2022). Victims of epistemic oppression are prone to develop the vices of *timidity* and *servility*. Meanwhile, those who benefit from oppression are prone to acquire the vices of *narcissism*, *vanity*, and *arrogance*. *Timidity* refers to behaviors of hiding, self-silencing, or testimonial smothering (Dotson 2011) because of fear of being exposed or perceived as intellectually inadequate. *Servility* refers to being too deferential to gain the approval of those in a dominant position, reinforcing and accepting one’s subordinate condition. *Vanity* is the disposition to seek the admiration and esteem of others, regardless of whether one is worthy of such praise. *Narcissists* have an excess of self-love and are convinced they are the standard of intellectual excellence. If they lack a particular feature, then it is not worth having. Finally, *arrogance* is an excess of self-confidence that leads agents to believe they are invulnerable and self-sufficient. This vice is characterized by a defensive attitude: i.e., the arrogant person deploys strategies to isolate themselves from criticism and react aggressively to anyone who questions their authority.

With this framework, I now examine the situation of many scientific communities located in subaltern regions. I aim to show that these communities often suffer oppression because of their location, identifying the institutional and material factors that shape and enable scientific practices that result in unfair conditions for scientific communities in LMICs and unfair gains for the already privileged.

14.4 Structural inequalities of global science

In this section, I describe some features of current global science that leave scientific communities in subaltern regions at a systematic disadvantage when participating in the production of knowledge. Nonetheless, it is important to highlight that these are only a few of the structural difficulties that scientific communities face.

14.4.1 Research agendas, funding, and extractivism

Researchers from LMICs are compelled to build ties with scientific communities from high-income countries to gain recognition and authority and increase visibility (Rodriguez-Medina et al. 2019). This is essential for pursuing any research due to (i) the scarcity of economic and material resources these communities frequently face and (ii) the hierarchical social structure of science. As it is well-known, researchers and institutions with higher prestige—primarily located in high-income countries—accumulate symbolic and material resources that, in turn, allow them to gain and secure even more prestige. The result of this phenomenon—the so-called Matthew Effect (Merton 1968)—is the same as any system of social stratification: “the rich get richer at the rate that [it] makes the poor become relatively poorer. Thus, centers of demonstrated scientific excellence are allocated far larger resources for investigation than centers which have yet to make their mark” (62). Given the current *project-based* funding model and high-competitive environments (Gläser and Laudel 2016), scientific communities often establish unequal partnerships and negotiate between their agendas and the interests of prestigious scientific communities, sometimes at the expense of their own needs. This has been described as *subordinate integration* (Feld and Kreimer 2020).

Against this background, we must consider the research and development (R&D) expenditure worldwide and the particular situation of LMICs (Fernández Pinto 2022). Taking the example of Latin America and the Caribbean, several aspects become salient. First, funding is mostly public. This contrasts with most high-income research-intensive countries where R&D expenditure and performance come mainly from private industry. Second, R&D expenditure is significantly low both in quantity and relative to their gross domestic product (GDP) compared to high-income countries (NSF 2022; OEI 2023). Thus, scientific communities are pushed to seek resources from “central” organizations that provide “development assistance” (e.g., the EU commission), foreign private industry (e.g., pharmaceutical companies), and (sometimes self-interested) “philanthropists” (e.g., Gates Foundation) (McGoey 2015; Vessuri 2017). Consequently, they may end up aligning their values with those of high-income countries and the private sector, neglecting research concerning problems that affect vulnerable communities in these regions (Yegros-Yegros et al. 2020).

Furthermore, scientific communities in high-income countries often engage in “North-South” collaborations to pursue their own interests through short-term projects (Gewin 2023; Schmidt and Pröpper 2017). In such partnerships, partners from LMICs do not participate in the research design, do not choose their tasks, do not manage resources, and have little power in deciding the use of the research results. Moreover, powerful actors, on many occasions, seek to extract biological data, specimens, and knowledge from local and indigenous communities (i.e., bioprospecting or biopiracy). Given the legislative gaps concerning the extraction of biological resources and weak intellectual property laws in some contexts, stakeholders from high-income countries manage to claim patents and commercialize these natural and epistemic resources, taking away such rights from local communities (Alcoff 2022; Gómez Lee and Roth Deubel 2023).

Likewise, pharmaceutical companies often undertake clinical trials with vulnerable populations from LMICs (Fernández Pinto 2019). With the supposedly laudable motive of doing research with subjects from underrepresented groups and offering treatment to communities that otherwise would not have access to it, companies take advantage of (i) the needs and the strong doctor-patient relations of these populations, (ii) patients’ lack of understanding of the treatment and its side-effects, and (iii) the large “treatment-naïve population” in the region. However, pharmaceutical products are later commercialized in the “Global North,” remain inaccessible in low-income countries, and do not respond to the needs of these vulnerable communities (Yegros-Yegros et al. 2020). Some even take advantage of allegedly weak regulations for performing research because they facilitate data gathering and procedures that could violate the rights of the most vulnerable (González-Santos and Saldaña-Tejeda 2023).

What I have just portrayed is a case of *epistemic powerlessness* and *epistemic exploitation*. Scientific communities in LMICs encounter difficulties in deciding their tasks in the division of cognitive labor, have less epistemic authority, the tasks they are assigned are often less valued or prevent them from fully exercising their capacities, and perform epistemic labor to fulfill others’ interests. In contrast, privileged scientific communities assign tasks to others and determine their own, thus having the means of developing and fully exercising their epistemic agency and determining the distribution of the benefits and burdens of global epistemic labor. This way, they profit and gain epistemic resources from scientific and non-academic communities in subaltern regions while setting back the latter’s interests.

This situation encourages three epistemic vices: *Servility* in scientists in a disadvantageous position and *vanity* and *arrogance* in scientists from high-income countries. The former are pushed to seek the approval and recognition of the latter to the point where they may be willing to sacrifice their needs and values, become reluctant to express disagreement, and agree to exploitative, unregulated, and extractivist practices because this may seem like the only way (or the most efficient way) of participating in knowledge production alongside prestigious scientific communities. Meanwhile,

privileged epistemic communities enjoy this deferential treatment, for it allows them to accumulate more epistemic and economic resources. This reinforces a sense of self-sufficiency, that is, the belief that they are needed but do not need anyone else, that they are righteous for “helping out” (even if they profit from it), and that their values are what all should be pursuing. The result is a lack of criticism and knowledge gaps in areas that could benefit the most vulnerable populations worldwide.

14.4.2 Academic migration

LMICs often do not have a strong institutional structure for research in comparison to research-intensive countries (Vessuri 2006). This motivates young scholars to pursue their graduate education in high-income countries, mostly in Europe and North America, and many choose not to return because of professional and personal reasons (Ramírez and Rodríguez-Medina 2024). This is not necessarily a bad thing. Still, we must consider that the migration of many talented people who decide to develop their careers abroad shrinks the already small pool of researchers in LMICs and increases the already big pool of researchers in high-income countries. In 2021, all of Latin America and the Caribbean concentrated 4% of researchers worldwide, while Africa, only 2%. Meanwhile, North America and Europe concentrated 43% of researchers, and Asia concentrated 49.4% (mostly from China and India) (OEI 2023). In addition, “South-North” migration may result in *ineffective epistemic diversity* (Fehr 2011): Scholars are encouraged to adapt their research agendas, conceptual frameworks, worldviews, background assumptions, and values to those of the powerful institutions where they develop their careers and obtain their degrees.

Furthermore, migration and internationalization carry significant economic and personal burdens (Ramírez and Rodríguez-Medina 2024). Besides the costs of traveling to the “top” international venues and institutions, visa requirements add a huge difficulty. Most scholars from LMICs have “precarious passports” (Asgarilaleh 2023), assume bigger migration expenses, and are more vulnerable to discrimination and racism (sometimes making it impossible for them to participate in epistemic exchanges).

These constraints result in *epistemic marginalization*. Communities in disadvantageous locations find constant difficulties in accessing resources and contributing to the production of scientific knowledge in mainstream and dominant venues. This, in turn, may result in timidity: Researchers could decide to avoid going through all the personal, emotional, and economic costs of participating in certain platforms. In such a context, scientific communities from privileged regions can develop arrogance: They remain shielded from possible criticism and reaffirm their alleged intellectual superiority and self-sufficiency. They may insist that they want their venues to be “diverse” (for they encourage the presence of members from the “Global South” within their institutions and events) but do not question the direction of academic mobility or think of other strategies for academic exchange. This can result in *epistemic exploitation*: The participation of scholars from geographically diverse origins in high-income institutions ends up favoring these resource-rich environments instead of the epistemic communities in disadvantageous regions.

14.4.3 Research evaluation and academic publishing

Research evaluation systems rely on metrics such as research output, citation, the type of venues where researchers circulate their results, and their recognition or visibility (Kulczycki 2023). This is generally measured in terms of the number of citations and published articles in high-impact journals, usually included in large bibliographic and citation indexes such as Scopus and Web of Science (Vessuri, Guédon, and Cetto 2014). Yet, these indexes are anything but neutral. On the one hand, historically, they began indexing journals mainly from Europe and North America. On the other hand, these

indexes were first conceived as tools for finding and identifying relevant literature and as indicators of knowledge circulation and networks, but not for assessing the quality of research (Voges 2023). Thus, belonging to these indexes is not a certification of the necessary “quality” to be part of the critical scientific exchange. On the contrary, it is a sign that one is already part of such an exchange. Consequently, it is not a surprise that these indexes have a language and geographical bias (De Albuquerque et al. 2020), marginalizing communities in LMICs.

Now, these are not the only venues for scientific knowledge circulation. Latin America, for example, has a long tradition of knowledge circuits thanks to the development of research associations, editorial industry, and bibliographic indexes during the second half of the twentieth century. Moreover, the region is a pioneer in Diamond Open Access and non-profit editorial management (Beigel 2019). Still, despite the efforts to consolidate regional platforms, there are factors that disincentivize scholars and institutions from using them. For instance, “non-mainstream” conferences and publishing venues are less valued in evaluation metrics and are often labeled “predatory” or “illegitimate.” Though it is undeniable that there are academic venues that seek profit instead of quality and have no or very weak peer-review processes, many have criticized the term “predatory,” pointing out the racist and geographic bias of this category (Krawczyk and Kulczycki 2021). Against this background, it is reasonable for scholars to avoid publishing in “unknown” platforms because of how this may affect the evaluation and perception of their research. Accordingly, they go for high-impact journals (many with a gold model of open access). Since OA is important for visibility, authors and institutions already in disadvantaged positions are required to pay article processing charges (APCs) ranging from 1000 USD to 4000 USD (Butler et al. 2023)—when paradoxically, high-quality regional journals would publish their work for free.⁵

These evaluation systems and quality standards could be described as cases of *cultural imperialism*. Privileged communities set themselves and their contexts as the standard of quality research and the legitimate places for knowledge production. Scientific communities in subaltern regions are consequently evaluated as epistemically inferior because they fail to fit these images of what “good” science is supposed to look like. This promotes *narcissism* among the privileged: They are the standard upon which others are judged. This shields them from criticism and renders them willfully ignorant, neglecting the existing venues and the epistemic resources that scientific communities in LMICs use and becoming unaware of possible disagreement, friction, and potential creativity. In addition, this situation encourages *timidity*: Scientific communities in subaltern regions develop practices of self-silencing and isolation because they become convinced that their work will not be positively valued in “top” venues (Mills and Inouye 2021). The consequences, at a macro-level, may result in hampering the existing venues where marginalized scientific communities resist and develop locally relevant epistemic resources, reducing criticism and innovation.

14.5 Conclusion: Taking geographic diversity seriously

The previous section illustrates how different forms of epistemic oppression are connected to the geographic location where scientists live and work. These locations play a crucial role in the creation of a hierarchical division of labor, decision-making power, and the distribution of epistemic resources, gains, and burdens. Many standard background conditions for doing research hinder the epistemic agency of scientific communities in LMICs while benefiting those in privileged locations. The consequence of these global asymmetries is a global science that mainly serves the values of the most privileged locations at the expense of the needs of the most vulnerable regions, and that ignores valuable contributions and criticism coming from non-hegemonic places.

In line with the literature on diversity, I believe that a promising avenue for managing and identifying these dominant values is to increase geographic diversity, that is, to encourage epistemic

exchanges among researchers located in different geographic locations and especially give uptake to those scientific communities in LMICs. Yet, this challenge cannot be addressed solely by moving researchers and diversifying institutions and venues in hegemonic regions, nor by fomenting acritical and unequal international partnerships, nor by primarily promoting transdisciplinary research with non-academic and “non-Western” communities from the “Global South.” Addressing all our attention to these goals may be part of the problem.

We should understand geographic diversity in terms of *positional difference* and not solely in terms of *cultural difference* and acknowledge that scientific communities occupy different positions of power because of their location. Accordingly, we must conceive strategies to counteract the factors involved in shaping the values of science and the participation of epistemic communities across the globe: Funding, venues for criticism, research evaluation, migration, and material and economic resources. Furthermore, we need to abandon the pervasive assumption that science is something proper of resource-rich contexts and examine how science is done in precarious conditions (without conceptualizing it as imitation, mimicry, or a derivative practice) (Dasgupta 2021).

With this at hand, we can begin to think about how to build *judicious connections* (Leonelli 2023) among geographically diverse scientific communities, i.e., connections that respect the epistemic autonomy and agency of scientific communities and institutions, especially those in low- and middle-income regions. In Leonelli’s (2023) words,

judicious connections bring a new experience of otherness, and the opportunity to modify one’s perspective and perception of the world. It is through the establishment of connections [...] that fruitful disagreements and frictions may come to light, and generate novel reactions and insight. (60)

Following Young’s (2004) proposal for reasoning about our political responsibility concerning global structural injustice, I suggest starting this effort with three parameters at hand: *Connection*, *power*, and *privilege*. Philosophers of science, science policymakers, and researchers who wish to increase diversity and address the value-ladenness of global science must take geographic location seriously and identify (i) how the actions of local/national scientific communities are connected to other scientific communities across the world, (ii) how much power one’s scientific community has within the global structure of science and identify which processes one could influence from one’s position, and (iii) how much one’s community or institution benefits from oppressive global structures. These parameters of reasoning could bring to light possible paths for collective action towards a more epistemically just global science.

Notes

- 1 Throughout the chapter, I use the terms “low- and middle-income” or “subaltern” to refer to locations comprised under the label “Global South.”
- 2 The distinction between “other” knowledges and “science” is a matter of debate (Whyte 2013). For the moment, I wish to highlight the silence around the *place* of (“Western”) science versus the constant mention of the place of all other epistemic practices.
- 3 There are a few works in this direction: Koskinen and Rolin (2021), Leonelli (2023), Dasgupta (2021), and Fernández Pinto (2019, 2022).
- 4 In this sense, epistemic oppression differs from Miranda Fricker’s (2007) concept of “epistemic injustice.”
- 5 Diamond Open Access is a publication model where journals publish articles in OA without charging authors APCs. In contrast, Gold Open Access consists of publishing texts in OA, though it is often conditional on authors paying APCs to the journal.

References

- Agnew, John. 2007. "Know-Where: Geographies of Knowledge of World Politics." *International Political Sociology* 1 (2): 138–48. <https://doi.org/10.1111/j.1749-5687.2007.00009.x>.
- Albuquerque, Ulysses Paulino, David Ludwig, Ivanilda Soares Feitosa, et al. 2021. "Integrating Traditional Ecological Knowledge into Academic Research at Local and Global Scales." *Regional Environmental Change* 21 (2): 45. <https://doi.org/10.1007/s10113-021-01774-2>.
- Alcoff, Linda Martin. 2022. "Extractivist Epistemologies." *Tapuya: Latin American Science, Technology and Society* 5 (1): 2127231. <https://doi.org/10.1080/25729861.2022.2127231>.
- Angus, Simon D., Kadir Atalay, Jonathan Newton, and David Ubilava. 2021. "Geographic Diversity in Economic Publishing." *Journal of Economic Behavior & Organization* 190 (October): 255–62. <https://doi.org/10.1016/j.jebo.2021.08.005>.
- Asgarilaleh, Tara. 2023. "On Being a 'Migrant Academic': Precarious Passports and Invisible Struggles." In *Migrant Academics' Narratives of Precarity and Resilience in Europe*, edited by Olga Burlyuk and Ladan Rahbari. Open Book Publishers. <https://doi.org/10.11647/OBP.0331>.
- Beigel, Fernanda. 2019. "Indicadores de Circulación: Una Perspectiva Multi-Escalar Para Medir La Producción Científico-Tecnológica Latinoamericana." *Ciencia, Tecnología y Política* 2 (3): 028. <https://doi.org/10.24215/26183188e028>.
- Berenstain, Nora. 2016. "Epistemic Exploitation." *Ergo, an Open Access Journal of Philosophy* 3 (22): 569–90. <https://doi.org/10.3998/ergo.12405314.0003.022>.
- Biddle, Justin B., and Anna Leuschner. 2015. "Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?" *European Journal for Philosophy of Science* 5 (3): 261–78. <https://doi.org/10.1007/s13194-014-0101-x>.
- Butler, Leigh-Ann, Lisa Matthias, Marc-André Simard, Philippe Mongeon, and Stefanie Haustein. 2023. "The Oligopoly's Shift to Open Access: How the Big Five Academic Publishers Profit from Article Processing Charges." *Quantitative Science Studies* 4 (4): 778–99. https://doi.org/10.1162/qss_a_00272.
- Dasgupta, Deepanwita. 2021. *Creativity from the Periphery: Trading Zones of Scientific Exchange in Colonial India*. University of Pittsburgh Press.
- De Albuquerque, Afonso, Thaianne Moreira De Oliveira, Marcelo Alves Dos Santos Junior, and Sofia Oliveira Firmo De Albuquerque. 2020. "Structural Limits to the De-Westernization of the Communication Field: The Editorial Board in Clarivate's JCR System." *Communication, Culture and Critique* 13 (2): 185–203. <https://doi.org/10.1093/ccc/tcaa015>.
- De Melo-Martín, Inmaculada, and Kristen Intemann. 2013. "Scientific Dissent and Public Policy: Is Targeting Dissent a Reasonable Way to Protect Sound Policy Decisions?" *EMBO Reports* 14 (3): 231–35. <https://doi.org/10.1038/embor.2013.8>.
- Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–57. <https://doi.org/10.1111/j.1527-2001.2011.01177.x>.
- Dotson, Kristie. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–38. <https://doi.org/10.1080/02691728.2013.782585>.
- Fehr, Carla. 2011. "What Is in It for Me? The Benefits of Diversity in Scientific Communities." In *Feminist Epistemology and Philosophy of Science*, edited by Heidi E. Grasswick. Springer.
- Fehr, Carla, and Janet Minji Jones. 2022. "Culture, Exploitation, and Epistemic Approaches to Diversity." *Synthese* 200: 465. <https://doi.org/10.1007/s11229-022-03787-8>.
- Feld, Adriana, and Pablo Kreimer. 2020. "Latinoamericanos En Proyectos Europeos: Asimetrías En La Cooperación Científica Internacional." *Ciencia, Tecnología y Política* 3 (4): 035. <https://doi.org/10.24215/26183188e035>.
- Fernández Pinto, Manuela. 2019. "Doubly Disadvantaged: On the Recruitment of Diverse Subjects for Clinical Trials in Latin America." *Tapuya: Latin American Science, Technology and Society* 2 (1): 391–407. <https://doi.org/10.1080/25729861.2018.1556237>.
- Fernández Pinto, Manuela. 2022. "Ciencia Comercial En América Latina: Análisis de Los Retos de La Financiación Privada de La Investigación." In *Conocimientos, Sociedades y Tecnologías En América Latina*, edited by Hebe Vessuri. Universidad de los Andes, Fondo de Cultura Económica.
- Fernández Pinto, Manuela, and Anna Leuschner. 2026. "Epistemic Intimidation and Illegitimate Value-Influences in Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Gewin, Virginia. 2023. "Pack up the Parachute: Why Global North–South Collaborations Need to Change." *Nature* 619: 885–87. <https://doi.org/10.1038/d41586-023-02313-1>.

- Gläser, Jochen, and Grit Laudel. 2016. "Governing Science: How Science Policy Shapes Research Content." *European Journal of Sociology* 57 (1): 117–68. <https://doi.org/10.1017/S0003975616000047>.
- Gómez Lee, Martha Isabel, and André-Noël Roth Deubel. 2023. "Traditional Knowledge Policy Co-Production in Colombia and Ecuador." *Tapuya: Latin American Science, Technology and Society* 6 (1): 2188107. <https://doi.org/10.1080/25729861.2023.2188107>.
- González-Santos, Sandra P., and Abril Saldaña-Tejeda. 2023. "'Hecho En México': A Media Analysis of the First MRT Baby." *Tapuya: Latin American Science, Technology and Society* 6 (1): 2245990. <https://doi.org/10.1080/25729861.2023.2245990>.
- Grosfoguel, Ramón. 2013. "The Structure of Knowledge in Westernized Universities: Epistemic Racism/Sexism and the Four Genocides/Epistemicides of the Long 16th Century." *Human Architecture: Journal of the Sociology of Self-Knowledge* 11 (1): 73–90.
- Harding, Sandra. 1993. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?" In *Feminist Epistemologies*, edited by Linda Alcoff and Elizabeth Potter. Routledge.
- Harding, Sandra. 2015. *Objectivity and Diversity: Another Logic of Scientific Research*. The University of Chicago Press.
- Harding, Sandra. 2018. "One Planet, Many Sciences." In *Constructing the Pluriverse: The Geopolitics of Knowledge*, edited by Bernd Reiter. Duke University Press.
- Intemann, Kristen. 2009. "Why Diversity Matters: Understanding and Applying the Diversity Component of the National Science Foundation's Broader Impacts Criterion." *Social Epistemology* 23 (3–4): 249–66. <https://doi.org/10.1080/02691720903364134>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Koskinen, Inkeri. 2022. "How Institutional Solutions Meant to Increase Diversity in Science Fail." *Synthese* 200 (6): 483. <https://doi.org/10.1007/s11229-022-03959-6>.
- Koskinen, Inkeri, and Kristina Rolin. 2021. "Structural Epistemic (in)Justice in Global Contexts." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo. Routledge.
- Kourany, Janet. 2010. *Philosophy of Science after Feminism*. Oxford University Press.
- Krawczyk, Franciszek, and Emanuel Kulczycki. 2021. "On the Geopolitics of Academic Publishing: The Mislocated Centers of Scholarly Communication." *Tapuya: Latin American Science, Technology and Society* 4 (1): 1984641. <https://doi.org/10.1080/25729861.2021.1984641>.
- Kulczycki, Emanuel. 2023. *The Evaluation Game: How Publication Metrics Shape Scholarly Communication*. Cambridge University Press.
- Lacey, Hugh. 2021. "Multi-Strategic Research and Traditional Saberes." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo. Routledge.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press.
- Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.
- Ludwig, David, Charbel N. El-Hani, Fabio Gatti, et al. 2024. "Transdisciplinary Philosophy of Science: Meeting the Challenge of Indigenous Expertise." *Philosophy of Science* 91 (5): 1221–31. <https://doi.org/10.1017/psa.2023.127>.
- McGoey, Linsey. 2015. *No Such Thing as a Free Gift: The Gates Foundation and the Price of Philanthropy*. Verso.
- Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159: 56–63.
- Mills, David, and Kelsey Inouye. 2021. "Problematising 'Predatory Publishing': A Systematic Review of Factors Shaping Publishing Motives, Decisions, and Experiences." *Learned Publishing* 34 (2): 89–104. <https://doi.org/10.1002/leap.1325>.
- Muldoon, Ryan. 2013. "Diversity and the Division of Cognitive Labor." *Philosophy Compass* 8 (2): 117–25. <https://doi.org/10.1111/phc3.12000>.
- NSF. 2022. "Science and Engineering Indicators: The State of US Science & Engineering 2022." National Science Foundation. <https://ncses.nsf.gov/pubs/nsb20221>.
- OEI. 2023. *El Estado de La Ciencia: Principales Indicadores de Ciencia y Tecnología Iberoamericanos/Interamericanos 2023*. Altuna Impresores.
- Page, Scott E. 2017. *The Diversity Bonus: How Great Teams Pay off in the Knowledge Economy*. Princeton University Press.
- Pesonen, Renne. 2022. "Argumentation, Cognition, and the Epistemic Benefits of Cognitive Diversity." *Synthese* 200 (4): 295. <https://doi.org/10.1007/s11229-022-03786-9>.
- Quijano, Anibal. 2000. "Colonialidad Del Poder, Eurocentrismo y América Latina." In *La Colonialidad Del Saber. Eurocentrismo y Ciencias Sociales: Perspectivas Latinoamericanas*, edited by Edgardo Lander. CLACSO.

- Ramírez, Viviana, and Leandro Rodríguez-Medina. 2024. "Well-Being and the Internationalisation of Academic Life: An Exploration from the Periphery." *Higher Education* 87: 1551–68. <https://doi.org/10.1007/s10734-023-01078-w>.
- Rodríguez-Medina, Leandro, Hugo Ferpozzi, Juan Layna, Emiliano Martín Valdez, and Pablo Kreimer. 2019. "International Ties at Peripheral Sites: Co-Producing Social Processes and Scientific Knowledge in Latin America." *Science as Culture* 28 (4): 562–88. <https://doi.org/10.1080/09505431.2019.1629409>.
- Schmidt, Laura, and Michael Pröpper. 2017. "Transdisciplinarity as a Real-World Challenge: A Case Study on a North–South Collaboration." *Sustainability Science* 12 (3): 365–79. <https://doi.org/10.1007/s11625-017-0430-8>.
- Steel, Daniel, and Karoline Paier. 2022. "Pro-Diversity Beliefs and the Diverse Person's Burden." *Synthese* 200 (5): 357. <https://doi.org/10.1007/s11229-022-03785-w>.
- Tanesini, Alessandra. 2022. "Intellectual Vices in Conditions of Oppression: The Turn to the Political in Virtue Epistemology." In *The Political Turn in Analytic Philosophy: Reflections on Social Injustice and Oppression*, edited by David Bordonaba Plou, Víctor Fernández Castro, and José Ramón Torices. De Gruyter.
- Vessuri, Hebe. 2006. "Academic Science in Twentieth-Century Latin America." In *Science in Latin America: A History*, edited by Juan José Saldaña. University of Texas Press.
- Vessuri, Hebe. 2017. "From Science as 'Development Assistance' to 'Global Philanthropy'." In *The Routledge Handbook of the Political Economy of Science*, edited by David Tyfield, Rebecca Lave, Samuel Randalls, and Charles Thorpe. Routledge.
- Vessuri, Hebe, Jean-Claude Guédon, and Ana María Cetto. 2014. "Excellence or Quality? Impact of the Current Competition Regime on Science and Scientific Publishing in Latin America and Its Implications for Development." *Current Sociology* 62 (5): 647–65. <https://doi.org/10.1177/0011392113512839>.
- Voges, Jonathan. 2023. "The Role of Bibliographic Indices for Knowledge Circulation." In *Routledge Handbook of Academic Knowledge Circulation*, edited by Wiebke Keim and Leandro Rodríguez-Medina. Routledge.
- Wallerstein, Immanuel Maurice. 2004. *World-Systems Analysis: An Introduction*. Duke University Press.
- Whyte, Kyle Powys. 2013. "On the Role of Traditional Ecological Knowledge as a Collaborative Concept: A Philosophical Study." *Ecological Processes* 2: 7. <https://doi.org/10.1186/2192-1709-2-7>.
- Wolf, Allison B. 2020. *Just Immigration in the Americas: A Feminist Account*. Rowman & Littlefield.
- Wylie, Alison. 2015. "A Plurality of Pluralisms: Collaborative Practice in Archaeology." In *Objectivity in Science: New Perspectives from Science and Technology Studies*, edited by Flavia Padovani, Alan Richardson, and Jonathan Y. Tsou. Springer.
- Yegros-Yegros, Alfredo, Wouter Van De Klippe, Maria Francisca Abad-Garcia, and Ismael Rafols. 2020. "Exploring Why Global Health Needs Are Unmet by Research Efforts: The Potential Influences of Geography, Industry and Publication Incentives." *Health Research Policy and Systems* 18 (1): 47. <https://doi.org/10.1186/s12961-020-00560-6>.
- Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton University Press.
- Young, Iris Marion. 2004. "Responsibility and Global Labor Justice." *The Journal of Political Philosophy* 12 (4): 365–88.

THE LIMITS OF DIVERSITY IN SCIENCE

The Case of Human Microbiome Research

Abigail Nieves Delgado, Jan Baedke, and Ann-Christin Fischer

15.1 Introduction

In recent years diversity emerged as a central value in society ([Sarkar 2010](#)). Prima facie, what comes to mind when thinking about diversity *in science* is the diversity or plurality of researchers and research subjects (e.g., patients). The current literature in philosophy of science emphasizes the importance of various forms of diversity and how these play crucial roles in advancing scientific progress ([Koskinen 2022](#); [Ludwig and Ruphy 2024](#); [Rolin et al. 2023](#)). In fact, as these studies show, the value of diversity can impact different stages and actions during research, such as data gathering or hypothesis building. In the following, we discuss diversity as “*data diversity*,” which points toward the idea to enrich databases by adding diverse samples from different groups and populations. Although, in theory, there is a high value in obtaining this diversity, there are several key issues and trade-offs in practice when aiming for it. To address these problems, we will draw particular attention to the problems of diversity in biomedical research, and more specifically in human microbiome research.

At least since the early 2000s, through the Human Genome Diversity Project and other national (meta-)genome projects, diversity has become key in genomics and related biomedical fields ([Bliss 2012](#); [Cavalli-Sforza 2005](#)). Collecting medically relevant “big data” about bodies (e.g., about genes, proteins, cells, tissues, DNA-methylation patterns), data about peoples’ health states, and even about lifestyles and cultural factors has become a driving force for pharmacological applications and health treatments. One particular field, in which diversity plays a crucial role, is human microbiome research. The human microbiome comprises bacteria, archaea, viruses, and fungi inhabiting various body sites, especially the gut. Microbiome composition impacts health in various ways and is linked to metabolic, autoimmune, and respiratory diseases, and maybe even to mental disorders ([Hou et al. 2022](#)). Environmental and cultural factors shape the microbiome, which connects health disparities to cultural practices. In this field, data is often collected from marginalized groups in non-Western settings (e.g., Indigenous communities) that are assumed to have a particularly species-rich microbiome. It is assumed that gathering this data about unique microbial taxa is highly relevant for revealing disease etiologies.

This movement toward data diversity is usually motivated by the attempt to gather (hopefully) *complete* datasets to allow better understanding of disease susceptibilities. It is expected to have both epistemic benefits, like addressing gaps and biases in data, and non-epistemic benefits, such as advancing health equity for everyone, including the investigated populations. In other words, data

diversity in biomedicine should serve a unified, global aim of health for all. However, as we will show, efforts to achieve more diverse and thus complete databases can lead to unintended harms. In microbiome research they trigger practices and forms of interactions with marginalized and minoritized communities that include racial stereotypes and othering, data extractivism, and sampling practices that reinforce colonial values.

In this chapter we will discuss the complexities of pursuing diversity in scientific databases. We argue that aiming for completeness for the purpose of achieving diversity can, in fact, undermine inclusivity and create ethical issues. In contrast, we advocate for prioritizing the local needs and aims of communities to create genuinely inclusive science and improve health outcomes while respecting cultural and ethical boundaries in research practices. In what follows, we, first, introduce the reader to current debates in philosophy of science about diversity and about the role diversity plays in microbiome research (Section 15.2). Then (Section 15.3), we discuss epistemic and non-epistemic trade-offs and problems that diversity-focused research faces. In Section 15.4, we redirect the focus toward a new understanding of diversity in the field, by highlighting local epistemic and non-epistemic aims of research and local needs of studied communities. Finally, we reflect more generally on the pitfalls of values like diversity and completeness and the global or local character of science (Section 15.5).

15.2 The value of diversity

When speaking of the value of diversity, this idea can be described along four dimensions: (a) Cognitive diversity, (b) demographic or social location diversity, (c) social value diversity (see Rolin et al. 2023), and (d) geographic diversity (Gutiérrez Valderrama 2026). Cognitive diversity refers to different intellectual perspectives such as expertise or problem-solving strategies present in a group. A group is considered demographically diverse if its members occupy different social locations, for example, when varying in gender, class, or “race” (Koskinen 2022). Social value diversity refers to differences in social and political values or is meant when members of the group represent different interest groups (Rolin et al. 2023). Last, geographic diversity refers to scientific communities living and working in different geographic locations (Gutiérrez Valderrama 2026). Although demographic diversity can enhance cognitive and social value diversity, it does not do so necessarily. As a consequence, all these forms of diversity should be targeted when trying to improve scientific practices. In such attempts, cognitive diversity is considered crucial to better science, as it introduces different perspectives and problem-solving strategies especially in complex problems (Pöyhönen 2017; Reijula and Kuorikoski 2021). Nevertheless, Sikimić (2023) points out that cognitive diversity is only beneficial if true efforts are made to include underrepresented viewpoints. If underrepresented groups or viewpoints are forced to assimilate into scientific reasoning and practices, the benefits of cognitive diversity diminish. Such cases where minoritized communities are only superficially included can lead to *epistemic exploitation* (Fehr and Jones 2022). Epistemic exploitation occurs when one party gains knowledge or benefits from another’s epistemic position, while simultaneously harming or undermining the interests of the latter. For example, this can happen if diverse researchers or communities are invited to participate in scientific investigations (e.g., in biomedical studies), but they are, in fact, exploited epistemically and do not benefit from their participation.

These different dimensions of diversity have long been debated in philosophy of science, under these or different labels. Since the 1980s feminist philosophers like Sandra Harding (1988, 1992) and Helen Longino (1987) have discussed *social diversity* and already in the 1970s scholars have argued in favor of *epistemic diversity* (Ludwig and Ruphy 2024). The former refers to understanding knowledge claims as socially situated and in relation to the cognitive, demographic, and social value diversity of the members of a scientific group, while the latter concerns the possibility of having different equally valid ways of explaining or knowing the world. Since these early works,

philosophers of science were increasingly unsatisfied with the limitations of unifying and generalist theoretical frameworks, particularly given the often fragmented and heterogeneous nature of actual scientific practice. Philosophers increasingly pointed out that scientific theories often resist reduction and that science is not built on a single fundamental method or ontology (Ludwig and Ruphy 2024; see also Dupré 1994; Feyerabend 1975). Both of these long-debated views of diversity—social and epistemic—will become important in the case of human microbiome research. However, in this and many other biomedically relevant fields, also another kind of diversity matters: *Data diversity*. It feeds on the idea that there are epistemic (and possibly societal) benefits if the data gathered stem from socially and (what is considered) biologically diverse sources—human populations with different lifestyles and cultural practices, genetic or epigenetic variations. This includes often marginalized groups (e.g., women, Black people) of allegedly “non-Western” communities (e.g., Indigenous communities).

15.2.1 Diversity in biomedical databases

Calls for social, epistemic, and data diversity are widespread in biomedicine, especially when linked to genetics and genomics. Today, to have diverse databases and to provide benefits for different populations are common mandates in biomedical research. However, these views of inclusive science were not always the status quo. For example, Steven Epstein (2007) argues that only since the 1980s an understanding of human diversity shaped scientific research and policies in the US. These changes came as a reaction to the “white, young male” standard that dominated medical thinking (see Dresser 1992): “we tended to want to reduce the human to that 60 kilogram white male, 35 years of age, and make that the normative standard—and have everything extrapolated from that tidy, neat mean, ‘the average American male’” (Epstein 2007, 2).

There was increasing evidence of the limitation of data generated only on one “type of body” and how this practice led to discrimination of minorities who failed to benefit from biomedical developments. However, approaches that addressed this challenge introduced new problems, as, right from the beginning, they adopted racial categories as the framework in which diversity should be conceptualized. This is not surprising, given that the US federal policy has collected data based on race and ethnicity labels since the late 1970s (Yanow 2003). This practice has shaped a global understanding of human diversity in biomedical research. In addition, this development led to what Epstein called “the rise of recruitmentology,” a growing emphasis on recruiting, sampling, measuring, and comparing minoritized groups in terms of sex, race, ethnicity, and age to understand differences in disease susceptibility and progression, response to treatments, among other alleged biological differences (Epstein 2008, 802). Maybe the most central consequences of this development was the renewed interest in racial differences as informative patterns of medical and biological differences between human groups and the reinscription of race in genetics (Fullwiley 2007) and post-genomics (Duster 2015).

The policies developed by institutions like the National Institute of Health (NIH) had worldwide effects. Catherine Bliss (2012) has shown that benchmark genomic projects, such as the International HapMap Project, were developed following federal race standards motivated by the general national priority of attending to health inequalities in the US. In short, such projects suggested that health inequalities can be addressed if there is biomedical data of each population or “race.” In other words, a complete database that accounts for previous data gaps about minoritized populations can help in solving health inequalities. This logic encouraged scientists worldwide to collect genetic samples from so-far neglected populations all around the world. In this development, the epistemic aim of *data completeness* became the operationalized counterpart to a very specific understanding of diversity and inclusion. In contrast, if certain groups are not included in an incomplete dataset, scientists speak

of *underrepresentation*. For example, Munung and colleagues (2022) highlight with respect to equitable genomics in Africa:

The post-genomics era promises a revolution characterized by precision medicine and the integration of genomics into almost every area of biomedical research. At the same time, there are concerns that if care is not taken, the genomics revolution may widen global inequities in science and health. In Africa, these concerns are primarily linked to the underrepresentation of African populations in genomics research. (411)

The same concern is widely shared in human microbiome research.

15.2.2 Diversity in microbiome research

In recent years the conceptual network between *data diversity*, *racial difference*, *inclusion*, *completeness*, and *underrepresentation* has been especially endorsed in microbiome research, where it has deeply shaped research aims and questions, and data collection practices. The collective genomes of microbiota (bacteria, archaea, viruses, and fungi) are known as the microbiome. Recent research has highlighted that human microbiome has strong links with health, as microbial composition and diversity correlates with various diseases. For example, it has been associated with metabolic disorders (Al-Assal et al. 2018), autoimmune conditions (Wen et al. 2008), and respiratory illnesses (Chunxi et al. 2020). The gut microbiome, in particular, influences the immune system (Macpherson and Harris 2004) and may even affect mental health (Clarke et al. 2013), including conditions like depression (Valles-Colomer et al. 2019). The microbiome is deeply shaped by environmental and cultural factors, such as diet, lifestyle, and cultural traditions (Durack and Lynch 2018; Nkera-Gutabara et al. 2022). This means that health differences between human populations are not only due to genetic differences between groups but due to cultural differences that are embodied through the microbiome (e.g., through nutritional habits).

Human groups are classified in microbiome research following traditional labels from genetics or genomics. Thus, microbiome researchers tend to apply a combination of racial categories (Caucasian, white, Black), ethnic names of groups (Yanomami, Hadza, Flemish), geographical labels (Latin Americans in the US, Otomí Indigenous in Mexico City, Mestizo women in the US, Asian), geo-political (Western, non-Western), anthropological (hunter-gatherer), and national categories (US Americans, Dutch, Japanese) in their studies (for more see Benezra 2020; De Wolfe et al. 2021; Maroney 2017; Nieves Delgado and Baedke 2021, 2024). These categories—especially race—often include a strong socio-cultural dimension, as nutritional habits, hygiene regimes, lifestyles, stressors, pollutions, economic differences, etc., are considered to affect the microbiome and thus health states of different groups. It is at this point where researchers' calls for data diversity become interlinked with demands for social and epistemic diversity, as scientists need to gather not only data from more "diverse bodies" but also data on the local cultural practices and knowledges that could result in different microbial patterns across groups. Including so-far neglected groups into datasets about microbial variation makes necessary, researchers argue, to sample groups that vary with respect to social practices that affect the microbiome and that might have health-relevant knowledge (e.g., traditional ecological knowledge) about life-style practices and foods that need to be gathered as well. Often this demand for data leads to comprehensive biological and, less common, ethnographic sampling practices.

The link between different forms of diversity is strengthened and justified through the rationale of incompleteness and underrepresentation. Following epistemic values like (data) completeness

should solve especially the representation gap between the industrialized northern countries and those of the Global South. For example, Allali and colleagues (2021) argue with respect to Africa:

With the advances in NGS (next generation sequencing) extending over a decade, it [...] is, however, disappointing that only 168 studies investigated the African human microbiome using this technology. Considering that Africa is made up of 54 independent countries with extremely diverse genetic backgrounds and cultures [...], Africa is under-represented in the global microbiome literature [...]. (45)

As these examples show, genomics and microbiome research share the aim of building (more) complete databases which are diverse and inclusive, both socially and epistemically. The epistemic objective to overcome existing data gaps is also legitimized through the assumption that more precise and accurate data will create more robust explanations or better understanding of the complex causal link between socio-cultural patterns, microbial states, and health states of human groups. In other words, in the field the non-epistemic values of diversity and inclusion in science are deeply inter-linked with epistemic values like completeness, accuracy, and precision. They together co-determine research aims, population identification, and sampling practices. As we will see now, however, the relationship between these two sets of values is characterized by deep tensions and trade-offs, both on the epistemic and non-epistemic side.

15.3 Epistemic and non-epistemic trade-offs

Data diversity in biomedicine and microbiome research does not come without a price. Scientists usually highlight the epistemic payoffs (e.g., more accurate explanations and predictions of individuals' health trajectories) and economic advantages (from pharmacological or probiotic treatments fine-tailored to specific populations to new medical procedures like fecal microbiota transplantation for cancer therapy) that are due to collecting more data from diverse communities. However, a closer look reveals that this assumption is illusive. The aim of data diversity comes with different epistemic and non-epistemic trade-offs and problems. Let us discuss these in detail.

15.3.1 Epistemic trade-offs

First, there is an *epistemic trade-off* when trying to establish more complete biomedical datasets. It refers to the value of precision. In recent years, with the advent of next-generation high-throughput DNA-sequencing approaches in metagenomics (which trace and sequence non-human DNA-fragments), it has become possible to collect and process more and more microbial data (Navas-Molina et al. 2017). This approach is considered necessary due to the sheer complexity of the human microbiome. We have nearly as many microbial cells in our bodies as human cells, and their functional roles are difficult to disentangle, due to, for example, microbial changes over seasons and depending on environmental factors (e.g., nutrition), and because at least some relations between microbial taxa and health states are functionally redundant (Tian et al. 2020; Yatsunenko et al. 2012). The field's big data approach and the move toward more diverse datasets enriched through samples from underrepresented groups should uncover the mechanistic details and causation of microbiome-linked diseases. However, while machine learning tools can help make sense of the vast amount of associations between microbial taxa obtained, this is not always the case. In fact, adding more data and more precision to the datasets does not necessarily lead to better explanations or understanding of microbiome-linked diseases.

The precision of an explanation refers to the level of specific detail in the *explanans* and *explanandum* (Baedke, Fábregas-Tejeda, and Vergara-Silva 2020; Ylikoski and Kuorikoski 2010). Mechanistic detail can stem from different biological levels, from molecular-genetic data to human population data. In the interventionist framework of causation (Woodward 2003), if an explanation carries more explanatorily relevant information, we can make more and (possibly) more important inferences to counterfactual scenarios and thus answer more so-called w-questions (“what-if-things-had-been-different-questions”). These questions explore how changes in *explanans* variables would affect the *explanandum*. Thus, added precision improves explanatory power and understanding. However, precision and mechanistic detail alone do not guarantee better explanation, as excessive or irrelevant detail can mislead (Nathan 2012; Salmon 1984) and precision faces trade-offs with other explanatory standards.

Precision is constrained by epistemic values like sensitivity and cognitive salience (Ylikoski and Kuorikoski 2010). Highly detailed explanations are more sensitive to changes in *explanans* or *explanandum*, making them less stable. In microbiome research, highly fine-grained causal dependencies between microbes and health states hold merely under a very small set of contextual factors (e.g., organismal or environmental states) and easily break under varying conditions. In fact, we currently see conflicting reports stating that specific metabolites and microorganisms can be both beneficial and pathogenic and that differences between sampled Western and non-Western populations sometimes play a role and sometimes not (Gupta, Paul, and Dutta 2017; Quin and Gibson 2020). For instance, *Lactobacillus iners* is linked to both vaginal health and dysbiosis (Petrova et al. 2017). To improve explanatory power, microbiome research may benefit from less detailed, more robust explanations (e.g., that do not try to trace relations between individual taxa and health states) and that allow for inferences across a wider range of counterfactual scenarios without breaking the dependency relation.

Furthermore, overly detailed explanations can reduce cognitive salience (Ylikoski and Kuorikoski 2010), making it harder to understand the relation between *explanans* and *explanandum*. This is a key challenge in microbiome research, where the complexity of the associations traced by explanations complicate inferences and developing biomedical interventions. In fact, while the field holds much promise to cure diseases, the tremendous amount of funding and strong support of the field in the last 15 years has not yet translated into many microbiome-based therapeutics and treatments.

These results tell us that while there is no direct epistemic harm in collecting a lot of diverse data (but see below on non-epistemic problems of this practice), we need to be thoughtful about how we use and analyze this data, because of the above mentioned epistemic trade-offs. If we are aiming for robust and less sensitive results and explanations of microbiome-caused health states in different groups, then we should be skeptical of whether all the data collected should actually be included in the analysis. In short, we might end up collecting a lot of data that ultimately may not be epistemically useful.

15.3.2 Non-epistemic trade-offs

A second set of trade-offs relevant when trying to develop more complete datasets concerns non-epistemic values, especially ethically problematic consequences of pursuing data diversity. These problems are (a) the strengthening of damaging stereotypes about minoritized groups and racialization processes, (b) the danger of promoting extractivist methodologies, and (c) the use of invasive methodologies in long-lasting sampling processes.

15.3.2.1 Completeness, stereotypes, and race

The call for inclusion in microbiome research has motivated sampling of (what are considered) non-Western populations for the sake of completing databases that mainly contain Western samples. The West/non-West divide is a pervasive juxtaposition in the field. It is used to describe populations (Xu

and Knight 2015), westernized diets and non-Western locations (Deering et al. 2019), differential abundance of specific species (mostly *Prevotella*) in the gut microbiome (Prasoodanan et al. 2021), westernization of the human gut microbiome resulting from immigration (Vangay et al. 2018), and the evolution of foraging lifestyles (as a synonym of non-US populations) to industrialized urban Western lifestyles (Gupta, Paul, and Dutta 2017). The comparative logic of these studies tends to oppose specific groups and populations against the West and its underlying normative ideas of civilized, developed, secular, modern, industrialized, and urban (Hall 2019, 142). This often leads to exoticism, stereotyping, and othering of non-Western populations and Indigenous groups, which are understood as ancestral, uncontacted, and pristine (Benezra 2020; Maroney 2017; Nieves Delgado and Baedke 2021, 2024; Núñez Casal 2024). Equally problematic is the use of racial categories in this field. Like in genetics and genomics, microbiome research tends to adopt the racial classifications used in the US as population descriptors (Fortenberry 2013). However, using these labels in other non-US contexts is especially inaccurate given the sensitivity of the microbiome to context and its seasonal variation (Smits et al. 2017).

15.3.2.2 Completeness and extractivism

The aim of completeness and associated methodologies leads to a second non-epistemic problem. The value of creating complete, diverse, and inclusive databases encourages researchers to expand their sampling practices to all populations around the world. When these practices fail to actually provide a payoff to those populations, microbiome research can become an extractivist practice. As Linda M. Alcoff (2022) explains, extractivism is a broad term that refers to the extraction of goods of many kinds (including knowledge; see Fehr and Jones 2022) by former colonial powers. In science, biopiracy and bioprospecting are good examples of such practices (Alcoff 2022, 2).

15.3.2.3 Completeness and invasive sampling

A third value trade-off that microbiome research faces concerns sampling methods and intensity of sampling. In contrast to genetics, the human microbiome is a highly dynamic ecosystem that changes over time and differs on different body sites. Any aspiration of data completeness needs to take into consideration this variation in populations. Thus, longitudinal studies trying to account for changes in the microbiome need to collect several samples over a certain time frame from a population under study. This sampling increases the invasiveness of these studies given the regularity with which the sampling has to be conducted and the fact that samples often have to be taken from many different body sites (e.g., samples of the vaginal microbiome or fecal samples). Taking microbial samples in this way can be perceived as a highly invasive and uncomfortable process (Mangola et al. 2022).

Given these epistemic and non-epistemic trade-offs, we argue that diversity in databases, operationalized as completeness, not only prohibits a better understanding of the health phenomena under study but often violates the ultimate purpose of inclusion of minoritized groups to benefit from biomedical research. In the following section, we present a potential solution that overcomes these problems of diversity research by focusing on local needs and aims.

15.4 Local diversities: Epistemic and non-epistemic dimensions

We argue that if we want to achieve the non-epistemic goal of diversity and inclusion for the sake of health for all, we need to abandon—somewhat paradoxically—the epistemic goal of completeness. In practice, this means that we question the shared global aim of including more populations by collecting their data to produce complete databases, and thus to generate results that will benefit all

populations. Instead, we suggest that research teams striving for inclusion need to define more clearly their local goals or aims¹ by promoting social diversity in their teams and attention to local diversities. In other words, a focus on local aims, local knowers, and local views on diversity can promote real inclusion and diversity in biomedicine. Therefore, we need to implement the following changes:

- a *Social diversity*: Research teams should include more often local communities as knowers. Local knowers are aware not only of practices, lifestyles, and forms of environmental interaction, but of the different ontological commitments and beliefs that are relevant for the local communities and their relation to the environment (Warbrick, Heke, and Breed 2023). Their inclusion in the research process can not only improve the adequacy of the research question and match them on local preferences, ontologies, and beliefs of the people under study but, more importantly, warrants their right to participate in science (Massimi 2025).
- b *Attention to local aims*: In addition to the epistemic gain of including local communities as knowers, locals are aware of the special needs of their communities in terms of nutrition, agriculture, diseases, environmental challenges (Mangola et al. 2022).

Taking into consideration (a) and (b) will set researchers in a better direction to generate inclusive results from their research.

- c *Revision of notions of diversity*: Scientific research teams should critically think about their research questions and the populations they want to sample to answer them. As shown before, the goal of achieving complete databases comes from an understanding of human diversity where race and ethnicity are understood as globally applicable descriptors of difference. While in certain contexts race and ethnicity may be important markers of difference, in other contexts other descriptors often turn out to be more relevant. In cases where researchers are trying to identify effects of environmental pollution, for example, proximity to a polluted site can be a more important descriptor than race and ethnicity. Researchers should try to select studied populations in a critical way and based on a reflection of suitable categories (see the NASEM 2023 report for initial guidelines). In addition, a better understanding of local notions of diversity and difference (e.g., how locals understand how they relate or differ from other neighboring populations) is crucial. Incorporating these local views can fruitfully contribute to finding adequate population descriptors for the research questions at hand.

We acknowledge that this shift in focus—from global diversity to local difference, knowledge, practices, and needs—brings challenges to scientific research teams. Having a focus on local diversity and aims and the abandonment of the strategy of data completeness can make it difficult for research teams to establish a fruitful scientific exchange with the broader scientific community which operates globally. In addition, the attention to local views of diversity will have the effect of generating datasets that may not be entirely compatible with the way diversity has been understood so far (mainly in terms of race, ethnicity, and nation). This can lead to unwanted effects like non-comparable results or incompatible data descriptions, which may stir tensions between local and global scientific communities. While this shift can generate results that more accurately fit specific populations, their particular collective identities, socio-cultural practices and idiosyncrasies, and environmental relations, and, thus, provide better knowledge and data to develop fine-tailored health solutions, it could also harm international scientific collaborations. Not least, since a new “locality of explanations” can go against established values like the generality (and generalizability) of scientific knowledge (see also Leonelli 2022, 2023).

At the same time, this transition may not be as difficult as expected. As we have shown in [Section 15.2](#), the understanding of human diversity in terms of race and ethnicity has a clear historical origin located in US science. In that regard, we are also dealing with a *local* understanding of diversity, merely masked as a universal or general one. This implies that also this understanding of diversity faces the same conceptual and practical challenges of comparability and generality of local frameworks mentioned before. This realization that scientists are already living in a complex world of local conceptualizations and value systems when it comes to studying different human populations makes it even more urgent to develop frameworks of epistemic pluralism, integration, and perspectivism ([Brigandt 2013](#); [Leonelli 2013](#); [Ludwig and El-Hani 2020](#); [Massimi 2022](#); [Mitchell 2003](#)). As a next crucial step, scientists and philosophers need to include more seriously nonscientific knowledge systems and foster transdisciplinary research ([De la Rosa et al. 2024](#)) also in biomedical fields. This will allow to equally value and reward different forms of knowledge production and to generate collaboration schemes without helicopter research ([Haelewaters et al. 2021](#)) and uneven power relations ([Smith 2018](#)).

15.5 Conclusions

Against what is commonly believed by philosophers of science, we have argued that diversity in science can be detrimental if implemented due to the wrong epistemic reasons and accompanied by methodologies and practices that harm marginalized groups (see also [Bluhm 2026](#)). This is often the case when “data diversity” becomes a goal, for example, in biomedical research. Data diversity refers to the collection of (socially, ethnically, racially) diverse human data which should provide epistemic benefits, such as overcoming gaps in dataset and improving understanding of diseases, and non-epistemic benefits like health for all. The pursuit of diversity of this kind is often motivated by the idea of completeness, like the completeness of datasets and knowledge about biological traits and causes of diseases. However, as we have argued, this completeness-driven view of diversity faces several epistemic and non-epistemic problems and trade-offs: It does not automatically lead to better explanations, as it produces more sensitive and less stable explanations, and brings about a number of ethical problems, as it easily triggers practices that fall prey to harmful stereotypes, extractivism, and invasive sampling. The case of human microbiome research exemplifies these challenges and pitfalls.

Against this detrimental usage of diversity in science, we advocate for shifting diversity-driven research away from completeness and toward prioritizing local knowers, aims, and diversity views while seriously engaging with epistemic diversity to promote truly inclusive science. However, foregrounding local epistemic and non-epistemic value systems in this way is not easy either. First, it can create tensions between global science and local approaches. For example, it requires fields like microbiome research to focus more on local contexts, which threatens prevailing goals like international comparability of results and translation of knowledge into global medical treatments. Second, it might create other epistemic trade-offs, as the new locality of explanations could clash with established values like generality of knowledge and virtues like extrapolation. Other non-epistemic trade-offs and practical challenges include, in the case of microbiome research, the complexity of integrating local and Indigenous knowledge systems about ecosystems and socio-cultural practices into biomedical knowledge ([Ojeda-Linares et al. 2024](#); [Warbrick, Heke, and Breed 2023](#)), the need for more thorough ethnological and anthropological and even sociological training of microbiologists, and the transition of the field toward an inter- and transdisciplinary framework in which different forms of knowledge can be exchanged by scientific and non-scientific actors.

This case shows that doing diversity right in science is not an easy task. But it is worth facing this challenge, as it can surely improve knowledge production in various ways and also the way scientists interact with populations under study. In any case, a half-hearted introduction of diversity, like in the case of data diversity, has no prospects in science.

Acknowledgments

We thank the editors for the constructive comments on earlier versions of this chapter. A.N.D. gratefully acknowledges the financial support from NWO (VI.Vidi.221F.014), J.B. from the German Research Foundation (DFG; project no. BA 5808/2–1), and A-C.F. from the Konrad Adenauer Foundation.

Note

- 1 In the literature on values in science, global aims are conceptualized as a way to avoid epistemic corruption. Lusk and Elliott (2024) introduce the “adequacy for purpose” framework to argue in favor of local aims in science.

References

- Al-Assal, Karina, Ana Cristina Martinez, Raquel Susana Torrinhas, Camila Siqueira Cardinelli, and Dan Linetzky Waitzberg. 2018. “Gut Microbiota and Obesity.” *Clinical Nutrition Experimental* 20: 60–64. <https://doi.org/10.1016/j.yclnex.2018.03.001>.
- Alcoff, Linda. M. 2022. “Extractivist Epistemologies.” *Tapuya: Latin American Science, Technology and Society* 5 (1): 2127231. <https://doi.org/10.1080/25729861.2022.2127231>.
- Allali, Imane, Regina E. Abotsi, Lemese Ah Tow, Lehana Thabane, Heather J. Zar, Nicola M. Mulder, Mark P. Nicol. 2021. “Human Microbiota Research in Africa: A Systematic Review Reveals Gaps and Priorities for Future Research.” *Microbiome* 9: 241. <https://doi.org/10.1186/s40168-021-01195-7>.
- Baedke, Jan, Alejandro Fábregas-Tejeda, and Francisco Vergara-Silva. 2020. “Does the Extended Evolutionary Synthesis Entail Extended Explanatory Power?” *Biology & Philosophy* 35: 20. <https://doi.org/10.1007/s10539-020-9736-5>.
- Benezra, Amber. 2020. “Race in the Microbiome.” *Science, Technology, & Human Values* 45 (5): 877–902. <https://doi.org/10.1177/0162243920911998>.
- Bliss, Catherine. 2012. *Race Decoded: The Genomic Fight for Social Justice*. Stanford University Press.
- Bluhm, Robyn. 2026. ““Every Cell Has a Sex”: Sex Essentialism in Biomedical Research.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Brigandt, Ingo. 2013. “Systems Biology and the Integration of Mechanistic Explanation and Mathematical Explanation.” *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4): 477–92. <https://doi.org/10.1016/j.shpsc.2013.06.002>.
- Cavalli-Sforza, Luigi L. 2005. “The Human Genome Diversity Project: Past, Present and Future.” *Nature Reviews Genetics* 6 (4): 333–40. <https://doi.org/10.1038/nrg1596>.
- Chunxi, Li, Liu Haiyue, Lin Yanxia, Pan Jianbing, and Su Jin. 2020. “The Gut Microbiota and Respiratory Diseases: New Evidence.” *Journal of Immunology Research* 2020: 2340670. <https://doi.org/10.1155/2020/2340670>.
- Clarke, Gerard, Sue Grenham, Paul Scully, et al. 2013. “The Microbiome-Gut-Brain Axis During Early Life Regulates the Hippocampal Serotonergic System in a Sex-Dependent Manner.” *Molecular Psychiatry* 18 (6): 666–73. <https://doi.org/10.1038/mp.2012.77>.
- Deering, Kane E., Amanda Devine, Therese A. O’Sullivan, Johnny Lo, Mary C. Boyce, and Claus T. Christophersen. 2019. “Characterizing the Composition of the Pediatric Gut Microbiome: A Systematic Review.” *Nutrients* 12 (1): 16. <https://pubmed.ncbi.nlm.nih.gov/31861722/>
- De La Rosa, Gabriela, Esther Milberg Muñiz, Charbel N. El-Hani, and David Ludwig. 2024. “Navigating between Promises and Realities of Transdisciplinary Research for Environmental Conservation.” *Human Ecology* 52: 697–708. <https://doi.org/10.1007/s10745-024-00520-6>.
- De Wolfe, Travis J., Mohammed Rafi Arefin, Amber Benezra, and Maria Rebolledo Gómez. 2021. “Chasing Ghosts: Race, Racism, and the Future of Microbiome Research.” *mSystems* 6 (5): e0060421. <https://doi.org/10.1128/msystems.00604-21>.
- Dresser, Rebecca. 1992. “Wanted Single, White Male for Medical Research.” *The Hastings Center Report* 22 (1): 24–29. <https://doi.org/10.2307/3562720>.
- Dupré, John. 1994. “Against Scientific Imperialism.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994 (2): 374–81. <https://doi.org/10.1086/psaprocbienmeetp.1994.2.192948>.
- Durack, Juliana, and Susan V. Lynch. 2018. “The Gut Microbiome: Relationships with Disease and Opportunities for Therapy.” *Journal of Experimental Medicine* 216 (1): 20–40. <https://doi.org/10.1084/jem.20180448>.

- Duster, Troy. 2015. "A Post-Genomic Surprise: The Molecular Reinscription of Race in Science, Law and Medicine." *British Journal of Sociology* 66 (1): 1–27. <https://doi.org/10.1111/1468-4446.12118>.
- Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. University of Chicago Press.
- Epstein, Steven. 2008. "The Rise of 'Recruitmentology': Clinical Research, Racial Knowledge, and the Politics of Inclusion and Difference." *Social Studies of Science* 38 (5): 801–32. <https://doi.org/10.1177/0306312708091930>.
- Fehr, Carla, and Janet Minji Jones. 2022. "Culture, Exploitation, and Epistemic Approaches to Diversity." *Synthese* 200 (6): 1–25. <https://doi.org/10.1007/s11229-022-03787-8>.
- Feyerabend, Paul. 1975. *Against Method*. 3rd ed. Verso.
- Fortenberry, J. Dennis. 2013. "The Uses of Race and Ethnicity in Human Microbiome Research." *Trends in Microbiology* 21 (4): 165–66. <https://doi.org/10.1016/j.tim.2013.01.001>.
- Fullwiley, Duana. 2007. "The Molecularization of Race: Institutionalizing Human Difference in Pharmacogenetics Practice." *Science as Culture* 16 (1): 1–30. <https://doi.org/10.1080/09505430601180847>.
- Gupta, Vinod K., Sandip Paul, and Chitra Dutta. 2017. "Geography, Ethnicity or Subsistence-Specific Variations in Human Microbiome Composition and Diversity." *Frontiers in Microbiology* 8. <https://doi.org/10.3389/fmicb.2017.01162>.
- Gutiérrez Valderrama, Juliana. 2026. "Values in Global Science and the Relevance of Geographic Diversity." In *Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Haelewaters, Danny, Tina A. Hoffman, and Adriana L. Romero-Olivares. 2021. "Ten Simple Rules for Global North Researchers to Stop Perpetuating Helicopter Research in the Global South." *PLOS Computational Biology* 17 (8): e1009277. <https://pmc.ncbi.nlm.nih.gov/articles/PMC8376010/>.
- Hall, Stuart. 2019. "The West and the Rest: Discourse and Power [1991]." In *Essential Essays, Volume 2: Identity and Diaspora*, edited by David Morley. Duke University Press.
- Harding, Sandra. 1988. *Feminism and Methodology: Social Science Issues*. Indiana University Press.
- Harding, Sandra. 1992. "Rethinking Standpoint Epistemology: What Is 'Strong Objectivity'?" *The Centennial Review* 36 (3): 437–70.
- Hou, Kaijan, Zhuo-Xun Wu, Xuan-Yu Chen, et al. 2022. "Microbiota in Health and Diseases." *Signal Transduction and Targeted Therapy* 7: 135. <https://doi.org/10.1038/s41392-022-00974-4>.
- Koskinen, Inkeri. 2022. "How Institutional Solutions Meant to Increase Diversity in Science Fail." *Synthese* 200 (6): 483. <https://doi.org/10.1007/s11229-022-03959-6>.
- Leonelli, Sabina. 2013. "Integrating Data to Acquire New Knowledge: Three Modes of Integration in Plant Science." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 (4 Pt A): 503–14. <https://doi.org/10.1016/j.shpsc.2013.03.020>.
- Leonelli, Sabina. 2022. "Open Science and Epistemic Diversity: Friends or Foes?" *Philosophy of Science* 89 (5): 991–1001. <https://doi.org/10.1017/psa.2022.45>.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press. <https://www.cambridge.org/core/elements/philosophy-of-open-science/0D049ECF635F3B676C03C6868873E406>.
- Longino, Helen. E. 1987. "Can There Be a Feminist Science?" *Hypatia* 2 (3): 51–64. <https://doi.org/10.1111/j.1527-2001.1987.tb01341.x>.
- Ludwig, David, and Charbel N. El-Hani. 2020. "Philosophy of Ethnobiology: Understanding Knowledge Integration and Its Limitations." *Journal of Ethnobiology* 40 (1): 3–20.
- Ludwig, David, and Stéphanie Ruphy. 2024. "Scientific Pluralism." In *The Stanford Encyclopedia of Philosophy (Fall 2024)*, edited by Edward N. Zalta and Uri Nodelman. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2024/entries/scientific-pluralism/>.
- Lusk, Greg, and Kevin C. Elliott. 2024. "Against Global Aims for Science: Values, Epistemic Priority, and a Local Aims Approach." *Synthese* 204 (2): 47. <https://doi.org/10.1007/s11229-024-04633-9>.
- Macpherson, Andrew J., and Nicola L. Harris. 2004. "Interactions between Commensal Intestinal Bacteria and the Immune System." *Nature Reviews Immunology* 4: 478–85. <https://doi.org/10.1038/nri1373>.
- Mangola, Shani Msafiri, Justin R. Lund, Stephanie L. Schnorr, and Alyssa N. Crittenden. 2022. "Ethical Microbiome Research with Indigenous Communities." *Nature Microbiology* 7: 749–56. <https://doi.org/10.1038/s41564-022-01116-w>.
- Maroney, Stephanie. 2017. "Reviving Colonial Science in Ancestral Microbiome Research." *Microbiosocial*, January 10. <https://microbiosocial.wordpress.com/2017/01/10/reviving-colonial-science-in-ancestral-microbiome-research/>.
- Massimi, Michaela. 2022. *Perspectival Realism*. Oxford University Press.
- Massimi, Michaela. 2025. "Local Knowledge and the Right to Participate in Science." *Philosophy of Science*: 1–25. <https://doi.org/10.1017/psa.2025.4>.

- Mitchell, Sandra. D. 2003. *Biological Complexity and Integrative Pluralism*. Cambridge University Press.
- Munung, Nchangwi Syntia, Jantina de Vries, and Bridget Pratt. 2022. "Towards Equitable Genomics Governance in Africa: Guiding Principles from Theories of Global Health Governance and the African Moral Theory of Ubuntu." *Bioethics* 36 (4): 411–22. <https://doi.org/10.1111/bioe.12995>.
- Nathan, Marco J. 2012. "The Varieties of Molecular Explanation." *Philosophy of Science* 79 (2): 233–54. <https://doi.org/10.1086/664791>.
- National Academy of Sciences, Engineering, and Medicine. 2023. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. The National Academies Press. <https://doi.org/10.17226/26902>.
- Navas-Molina, Jose A., Embriette R. Hyde, Jon G. Sanders, and Rob Knight. 2017. "The Microbiome and Big Data." *Current Opinion in Systems Biology* 4: 92–96. <https://doi.org/10.1016/j.coisb.2017.07.003>.
- Nieves Delgado, Abigail, and Jan Baedke. 2021. "Does the Human Microbiome Tell Us Something about Race?" *Humanities and Social Sciences Communications* 8: 97. <https://doi.org/10.1057/s41599-021-00772-3>.
- Nieves Delgado, Abigail, and Jan Baedke. 2024. "How to Eliminate Race from Human Microbiome Research." *Argumenta*: 1–14. <https://doi.org/10.14275/2465-2334/20240.nie>.
- Nkera-Gutabara, Claudine K., Robyn Kerr, Janine Scholefield, Scott Hazelhurst, and Jerolen Naidoo. 2022. "Microbiomics: The Next Pillar of Precision Medicine and Its Role in African Healthcare." *Frontiers in Genetics* 13: 869610. <https://doi.org/10.3389/fgene.2022.869610>.
- Núñez Casal, Andrea. 2024. "Race and Indigeneity in Human Microbiome Science: Microbiomisation and the Historiality of Otherness." *History and Philosophy of the Life Sciences* 46 (17). <https://doi.org/10.1007/s40656-024-00614-w>.
- Ojeda-Linares, César, Alejandro Casas, Tania González-Rivadeneira, and Gary P. Nabhan. 2024. "The Dawn of Ethnomicrobiology: An Interdisciplinary Research Field on Interactions between Humans and Microorganisms." *Journal of Ethnobiology and Ethnomedicine* 20 (86). <https://doi.org/10.1186/s13002-024-00725-5>.
- Petrova, Mariya I., Gregor Reid, Mario Vaneechoutte, and Sarah Lebeer. 2017. "*Lactobacillus Iners*: Friend or Foe?" *Trends in Microbiology* 25 (3): 182–91. <https://doi.org/10.1016/j.tim.2016.11.007>.
- Pöyhönen, Samuli. 2017. "Value of Cognitive Diversity in Science." *Synthese* 194 (11): 4519–40. <https://doi.org/10.1007/s11229-016-1147-4>.
- Prasoodanan, Vishnu, Ashok K. Sharma, Shruti Mahajan, et al. 2021. "Western and Non-Western Gut Microbiomes Reveal New Roles of *Prevotella* in Carbohydrate Metabolism and Mouth-Gut Axis." *Biofilms and Microbiomes* 7: 77. <https://doi.org/10.1038/s41522-021-00248-x>.
- Quin, Candice, and Deanna L. Gibson. 2020. "Human Behavior, Not Race or Geography, Is the Strongest Predictor of Microbial Succession in the Gut Bacteriome of Infants." *Gut Microbes* 11 (5): 1143–71. <https://doi.org/10.1080/19490976.2020.1736973>.
- Reijula, Samuli, and Jaakko Kuorikoski. 2021. "The Diversity-Ability Trade-Off in Scientific Problem Solving." *Philosophy of Science* 88 (5): 894–905. <https://doi.org/10.1086/714938>.
- Rolin, Kristina, Inkeri Koskinen, Jaakko Kuorikoski, and Samuli Reijula. 2023. "Social and Cognitive Diversity in Science: Introduction." *Synthese* 202 (2): 36. <https://doi.org/10.1007/s11229-023-04261-9>.
- Salmon, Wesley C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press.
- Sarkar, Sahotra. 2010. "Diversity: A Philosophical Perspective." *Diversity* 2 (1): 127–41. <https://doi.org/10.3390/d2010127>.
- Sikimić, Vlasta. 2023. "Epistemic Inclusion as the Key to Benefiting from Cognitive Diversity in Science." *Social Epistemology* 37 (6): 753–65. <https://doi.org/10.1080/02691728.2023.2258831>.
- Smith, James. 2018. "Parasitic and Parachute Research in Global Health." *The Lancet Global Health* 6 (8): e838.
- Smits, Samuel A., Jeff Leach, Erica D. Sonnenburg, et al. 2017. "Seasonal Cycling in the Gut Microbiome of the Hadza Hunter-Gatherers of Tanzania." *Science* 357: 802–6. <https://doi.org/10.1126/science.aan4834>.
- Tian, Liang, Xu-Wen Wang, Ang-Kun Wu, et al. 2020. "Deciphering Functional Redundancy in the Human Microbiome." *Nature Communications* 11: 6217. <https://doi.org/10.1038/s41467-020-19940-1>.
- Valles-Colomer, Mireia, Gwen Falony, Youssef Darzi, et al. 2019. "The Neuroactive Potential of the Human Gut Microbiota in Quality of Life and Depression." *Nature Microbiology* 4 (4): 623–32. <https://doi.org/10.1038/s41564-018-0337-x>.
- Vangay, Pajau, Abigail J. Johnson, Tonya L. Ward, et al. 2018. "US Immigration Westernizes the Human Gut Microbiome." *Cell* 175 (4): 962–72.e10.
- Warbrick, Isaac, Deborah Heke, and Martin Breed. 2023. "Indigenous Knowledge and the Microbiome—Bridging the Disconnect between Colonized Places, Peoples, and the Unseen Influences that Shape Our Health and Well-Being." *mSystems* 8: e00875–22. <https://doi.org/10.1128/msystems.00875-22>.

- Wen, Li, Ruth E. Ley, Pavel Y. Volchkov, et al. 2008. "Innate Immunity and Intestinal Microbiota in the Development of Type 1 Diabetes." *Nature* 455 (7216): 1109–13. <https://doi.org/10.1038/nature07336>.
- Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press. <https://doi.org/10.1093/0195155270.001.0001>.
- Xu, Zhenjiang, and Rob Knight. 2015. "Dietary Effects on Human Gut Microbiome Diversity." *The British Journal of Nutrition* 113: 1–5. <https://doi.org/10.1017/S0007114514004127>.
- Yanow, Dvora. 2003. *Constructing Race and Ethnicity in America: Category-Making in Public Policy and Administration*. Routledge. <https://doi.org/10.4324/9781315705460>.
- Yatsunenko, Tanya, Federico E. Rey, Mark J. Manary, et al. 2012. "Human Gut Microbiome Viewed Across Age and Geography." *Nature* 486: 222–27. <https://doi.org/10.1038/nature11053>.
- Ylikoski, Petri, and Jaakko Kuorikoski. 2010. "Dissecting Explanatory Power." *Philosophical Studies* 148 (2): 201–19. <https://doi.org/10.1007/s11098-008-9324-z>.

TRANSPARENCY IN SCIENCE

Stephen John

Imagine a group of researchers claim to have developed a new paradigm for cancer research. A legislator is uncertain whether to fund them, so insists on access to their lab. This allows her to check that the scientists aren't frauds, fabricating their experiments. Beyond that, however, it's not clear who gains from this radical transparency. The legislator's presence might limit the scientists, by inhibiting them from expressing doubts or trying out new ideas. Nor does unfettered access necessarily help the legislator make her decision, because she may be unable to understand what she sees. Of course, she could go beyond merely observing the scientists and ask them to explain themselves. However, this demand imposes even greater costs on the researchers, and it's not clear she has solved her problem; if she didn't trust the researchers' claims about their results, why trust their claims about their processes? A more sensible strategy, then, might be to give access to independent experts and ask their judgment. But now the legislator has a new problem: Should she demand the independent experts are transparent about how they do their work? If so

It is a commonplace that science ought to be more transparent. However, as my hypothetical example shows, such calls are both vague—transparency can come in many forms—and contestable—transparency can have costs as well as benefits. More generally, transparency stands in a complex relationship to trust: Calls for greater transparency are often motivated by worries about trustworthiness, but transparency may simply change whom we trust, rather than remove the need for trust. This chapter provides an overview of the roles and consequences of transparency in science, paying particular attention to debates over non-epistemic values in science. In [Section 16.1](#), I map out some reasons for, forms of, and dangers of transparency. In [Section 16.2](#), I focus on the relationship between transparency and concerns about non-epistemic values in science. Finally, in [Section 16.3](#), I explore the relationship between transparency, trust, and non-epistemic values.

Before going on, note that, although I focus on the costs and problems of transparency, this is not intended as a blanket condemnation. There are many cases where transparency regimes have been hugely beneficial, and many current practices would be improved by greater transparency. I adopt a skeptical tone because transparency is sometimes treated as a panacea for problems of scientific misconduct or tensions between science and society. This optimism downplays not only the costs of transparency but, in a way, its positive power; any genuinely effective tool will always be dangerous.

16.1 Transparency: Methods, aims, and costs

The general concept of transparency is to make some hidden or inaccessible process more visible (Heald 2022). For example, publishing the minutes of meetings of a scientific expert advice committee would increase transparency. Calls for greater transparency are complex, however, because notions of “visibility” are hard to unpack. First, there are different ways of being transparent: For example, rather than publish minutes, we could televise committee meetings.¹ Second, a process can be made transparent to some audiences but not others; for example, circulating minutes inside the civil service increases transparency for the rest of the governmental apparatus, but not the general public. Third, and most importantly, there is a gap between what Kevin Elliott calls being “visible” and being “legible” (Elliott 2022); broadcasting a committee’s meetings on television makes its workings “visible” to all, but, if deliberations are framed in a technical vocabulary, the process will remain “illegible” to many. These issues are inter-related: For example, if we want the committee’s decision-making to be “legible” to the public, making their work accessible to intermediaries, who can explain what is happening, may be more effective than simply making their deliberations visible to all.

To decide what kinds of transparency we want, it is useful to think about what transparency is *for*. Broadly, we can distinguish two sets of problems which greater transparency is supposed to resolve.

The first set clusters around “epistemic bad practice.” There are various ways in which scientists can go wrong in a distinctively epistemic sense; these range from downright fraud to more complicated “questionable research practices,” as, for example, when the pharmaceutical industry consigns “unfavorable” studies to the desk drawer (Every-Palmer and Howick 2014). One possible function of transparency is to tackle these problems. If scientists are transparent, it may be easier to spot when they have breached epistemic norms (or, at least, to apportion our epistemic credence more carefully), and the threat of exposure may dis-incentivize epistemic bad practice in the first place.²

The second set of problems clusters around “problematic values” concerns. As this volume explores, scientific practice is, in a variety of ways, shaped by non-epistemic values. We might worry, then, that science is shaped by the *wrong* values, or, at least, that audiences should know *which* values shape science. A second possible function of transparency is to uncover, and perhaps shape, these implicit values.

Note that, in many discussions, demands for transparency are often linked to demands for “open science,” involving greater sharing of data and techniques between researchers (Leonelli 2023). The precise relationship between transparency and open science demands is, however, complicated. In this chapter, then, I will focus on cases where science is “opened up” to an *outside* community, rather than the primary concern of “open science” advocates, of greater openness *within* scientific communities.

The broad goals of detecting epistemic bad practice and detecting problematic values are uncontroversial. In turn, identifying these goals can help us identify which sorts of transparency are valuable: For example, simply publishing highly edited minutes seems a bad way of promoting either. Still, moving from these goals to more concrete recommendations is difficult, for several reasons.

First, the effectiveness of transparency strategies cannot be settled *a-priori* or by common sense. For example, you might think it is obvious that asking researchers to declare “conflicts of interests” (a paradigm transparency strategy) is a good way of handling epistemic bad practice. Unfortunately, some research suggests that asking authors to declare their “conflicts of interest” actually *increases* the chance they will allow their interests to influence their research, by activating a kind of “buyer beware” reflex (de Melo-Martín and Intemann 2009). In this case, transparency may worsen epistemic bad practice! Similarly, publishing minutes of FDA meetings debating drug regulation may not allow us to understand the values implicit in those decisions if patient advocacy groups are, in fact, little more than fronts for corporate values (Holman and Geislar 2018).

Second, mere transparency cannot achieve anything by itself but must be considered against a social and institutional background. Consider a political analogy: Simply revealing that policymakers are corrupt won't change the behavior of shameless politicians. Rather, we may need mechanisms to bring them into line (Kogelmann 2021). Similarly, simply knowing that the pharmaceutical industry engages in questionable research practices doesn't stop these practices. Rather, we need further mechanisms which translate transparency into (a simulation of) public-mindedness. We cannot assess transparency measures without investigating incentives and institutions.

Third, while both goals are valuable, it is not clear that they are both best served by the *same* strategies. Non-epistemic values can be inculcated during basic training in a research field, such that these assumptions seem "obvious" to researchers; for example, as Elisabeth Lloyd has shown, androcentrism is a "systematic bias" in research into female sexuality (Lloyd 2006). Making research practices accessible to other trained experts may help to avoid epistemic bad practice, but nothing to challenge these non-epistemic pre-suppositions. For that task, we may need a different, more demanding, form of transparency, where we open up a field to diverse "outsider" eyes and voices. However, we then quickly run into problems of identifying who should have standing in an epistemic community (Koskinen 2017). In this sort of case, you may think that, as long as we are careful in our institutional design, greater public involvement will help improve *both* the epistemic and ethical quality of research, because problematic values may lead us away from a nuanced understanding of reality. In the rest of this section, I will argue against this elision. Rather, I claim, *even if* transparency succeeds in the goal of elucidating values, it can have epistemic costs. (Of course, transparency can also have other costs—for example, the financial costs of paying someone to pre-register trials—but I'll place those to one side.) The next section then considers more closely the links between transparency and identifying (problematic) non-epistemic values.

One possible epistemic cost of transparency is for non-scientists themselves. To understand this phenomenon, consider the fact that a much larger proportion of the US population claims to trust science than to believe climate change is occurring, despite awareness of the overwhelming consensus within climate science on the reality of anthropogenic climate change (Kovaka 2021). Karen Kovaka has argued that this "pro-science science denialism" emerges because of a gap between non-scientists' *beliefs* about good practice and *actual* good practice (Kovaka 2021). For example, imagine the public holds a "theory of science" according to which scientists must never be guided by their own values; they then learn that climate scientists breach this norm; as a result, they mistrust climate scientists. This can happen even if, as many contributors to this volume argue, good science is and ought to be value-laden.

Building on Kovaka's analysis, we can identify two possible epistemic costs of transparency. First, and most obviously, there might be costs *for the audiences*, who end up mistrusting scientists who are, in fact, trustworthy. For example, consider the mass release of emails from the Climate Research Unit (CRU) at the University of East Anglia. This was, in one sense, an exercise in radical transparency (John 2018). Climate skeptics argued that these emails showed that climate scientists were untrustworthy, even though the procedures they described (for example, disregarding some recalcitrant data in model-building) were epistemically justifiable. I have argued that, in this case, a "folk philosophy of science" may have led transparency to undermine epistemically well-grounded trust (John 2018). In effect, if we start from a situation where publics do trust scientists, transparency can lead to an epistemic loss for audiences. (Section 16.3 takes up the case where publics do not trust scientists.)

As well as having epistemic costs for the audience, transparency plus false beliefs about "good science" might lead to costs *for the quality of the research itself*. This can happen if researchers feel compelled to adopt sub-optimal epistemic strategies to ensure that their audiences continue to trust them. In a wide-ranging paper, C. Thi Nguyen (2022) has explored this

possibility, suggesting that transparency may require researchers to limit themselves to standardized procedures, easily understandable by, and amenable to, the public. However, much of their expertise is grounded in tacit knowledge, which is, by its nature, esoteric. By conforming to audiences' expectations that they reason in ways which audiences can understand, experts undermine their own expertise.

Clearly, these epistemic costs of transparency are contingent on the public holding a false philosophy of science, where science is a value-free, neat, algorithmic process which can be grasped by outsiders. So, it might seem that they can be avoided; for example, rather than a "data dump" of the CRU emails, it would have been better to ensure transparency through a process which allowed climate scientists to contextualize their workings (Hulme and Ravetz 2009), or that we should do a better job of educating the public about science more generally (Wilholt 2023).

However, there are three reasons to worry that such strategies will not counter the epistemic costs of transparency. First, in cases where we have to contextualize our actions, we are automatically at a rhetorical disadvantage; "I know it looks bad, but I can explain" is a hard message to get across! Second, to engage in public education, we must formulate the "correct" theory of science to communicate. Unfortunately, if there is consensus in philosophy of science, it is that there is no such thing as the "scientific method," rather we may need "correct" theories on a field-by-field basis. Furthermore, there remain large disagreements over the proper account of various fields (consider debates over the proper role of models in climate science or of RCTs in biomedicine). It is not obvious what the content of "the" theory of science ought to be; even if it were, the content is likely to be highly complex, and, hence, hard to communicate. Third, there are powerful, well-resourced social actors who have a vested interest in promulgating false (or, at least, highly simplified) "theories of science" (Oreskes and Conway 2010); we are not just saying that "this looks bad, but I can explain" and trying to communicate a complex picture, but doing so in an environment where others offer a far simpler, seductive picture: "it looks bad because it is bad, because it breaches the scientific method." Note that none of this is to deny that, eventually, education might succeed and obviate the costs of transparency. It is, however, to express some pessimism on this issue.

These concerns may point to a more fundamental problem with reconciling the two goals of improving epistemic practice and avoiding problematic values. As Nguyen stresses, a key feature of transparency is that it encourages the scrutinized to frame their discussions in ways which are well aligned with the values and understandings of the scrutinizers (Nguyen 2022). In political philosophy, the value of transparency is often framed in similar terms: For example, some hold that making policy deliberations more transparent forces policymakers to cleave to the demands of "public reason."³ Moving from the political context to the scientific context, we can understand why transparency may seem a plausible response to concerns about "value-laden" science; it drags scientists into the realm of public values (see Section 16.2). However, it is precisely the fact that transparency requires reasoning processes to become "public" which can lead to epistemic losses. A striking example of this phenomenon may be found in Ted Porter's work, exploring how demands that regulatory sciences be based on easily accessible "public" standards of justification can come into tension with employing the most reliable techniques on a case-by-case basis (Porter 1995, chap. 4). Of course, these remarks are highly schematic. Still, it is worth considering that "making science public" can be both *politically* valuable and *epistemically* harmful.

16.2 Non-epistemic values in science and transparency

The previous section argued that, even if transparency allows us to determine which values structure science, it might have epistemic costs. That is to say, there is a trade-off. But to think through that trade-off, we need to ask more about the link between transparency and values in

science. Why might it be important to know which values structure science? And what, precisely, is the relationship between this goal and calls for transparency? This section takes up these challenges.

In an influential paper, Gregor Betz briefly suggests an ethical argument for the “Value-Free Ideal” for science: That value-laden science is incompatible with respect for autonomy (Betz 2013). Betz’s suggestion seems to follow from two plausible claims. The first claim concerns the epistemology of testimony: Non-experts ought to defer to scientific experts’ factual claims about the world. In support of this claim, consider the anger many feel at politicians who refuse to accept scientists’ claims that the climate is warming. The second claim is ethical: We ought to respect others’ autonomy, understood as their ability to follow their own values. If scientists’ claims are value-laden, then these two plausible claims seem to conflict, because, in deferring to experts’ value-laden claims, non-experts risk (implicitly) endorsing ethical commitments they do not share.

Unfortunately, Betz’s argument faces an “ought implies can” problem: It is unclear that scientific justification *can* be value-free. As an example, consider what Anna Alexandrova calls “mixed claims”; for example, “beyond a particular level, increases in income do not improve well-being” (Alexandrova 2017). On the one hand, such claims can clearly be established or tested through scientific methods. On the other hand, the concept of “well-being” is value-laden; when people disagree over whether well-being is a matter of desire-satisfaction or objectively valuable functioning, their disagreements are ethical. So, if “mixed claims” have a proper place in science, “Value-Free Science” is an illusion.

Still, that does not mean that we escape Betz’s worries about autonomy. Indeed, Alexandrova herself suggests that a key problem in operationalizing well-being is that we risk “imposing” values on our audiences (Alexandrova 2017). If, for example, scientists conceive of well-being as desire-satisfaction and their claims are used for policy, this apparently forces a particular view of well-being on those who think in fundamentally different terms.

We can, then, understand many calls for transparency as stemming, ultimately, from a wish to preserve autonomy while recognizing that science is “value-laden.” For example, McKaughan and Elliott (2013) argue that making the role of non-epistemic values in research transparent is important, because it provides audiences with three options: They can demand the research is rerun with different values in place; if this is not possible or appropriate, they can try to assess whether the research results would be different had different values been in place; if this is impossible or inappropriate, they could just decide to disregard this research. In turn, they suggest that all three options help ensure that science communication does not undermine autonomy.

However, there is a problem in spelling out these links between autonomy and transparency. Prototypical concerns about violations of autonomy arise in cases where some agent *intentionally* or *deliberately* disregards another’s values, as, for example, when a paternalistic doctor overrules a patient’s stated wishes. Of course, scientists do sometimes deliberately employ controversial non-epistemic value judgments in researching, justifying, or communicating claims. However, most concerns about “non-epistemic” values in scientific justification are more complicated. We might, for example, be concerned that use of a particular “*p* value” in statistical testing is problematic because it risks under-regulation (Douglas 2000). However, it is a mistake to think that scientists who use that statistical test must have done so through an analysis of the costs and benefits of under- and over-regulation. Rather, they may just have been “doing what they always do.” When we say their choice of significance level was “value-laden,” we don’t mean that the choice *was* guided by a conscious value judgment, but that the choice can be criticized on non-epistemic grounds, because it is “as if” it reflected a value judgment (John 2021, chap. 2).⁴ Indeed, even in the obviously “value-laden” area of well-being, Alexandrova suggests that imposition is often the result of “inattention,” a failure to reflect on one’s own biases and assumptions (Alexandrova 2017).

The fact that science might be “value-laden” without scientists consciously making evaluative choices has important implications for transparency: Requiring that scientists’ own thought processes are visible will not necessarily make non-epistemic values visible. Simply asking economists how they ended up using a particular construct to measure well-being will not necessarily tell us which values are implicit in that construct; seeing regulatory scientists use a particular “*p* value” will not reveal the implicit value judgments encoded in that “choice.” This is not because scientists are *trying* to hide something. Instead, it is because they don’t have anything to hide.

As such, authors interested in the role of non-epistemic values in science typically suggest that we need more complex models of transparency, which allow us to excavate implicit values. Typically, these models involve two inter-locking features. First, they demand more than mere visibility, but that scientists enter into debate and discussion, which allows for probing of apparently “routine” choices (Alexandrova 2017). Second, because of demands about “groupthink” or “inattention,” they demand that debate involves a wide variety of diverse voices, well beyond the “expert” community (Elliott 2022).⁵ If the influence of values in science goes beyond conscious choice, the appeal of these interactive approaches is obvious. However, they face problems. Some of these problems are about the assumption that we can uncover implicit values in research. However, I will focus on a second set of issues:

even if critical dialogue can establish both which non-epistemic values structure research and that they are not aligned with the audience’s values, what happens next?

As noted above, McKaughan and Elliott suggest three responses. The first is that the experts should redo their work in line with the audience’s values. However, it is not clear that such demands are always viable: For example, if no-one has ever thought to operationalize well-being in the audience’s preferred way, it may be impossible to repeat a study using the audience’s preferred account. These problems become even more complicated if, as we discussed in Section 16.1, entire research fields are structured by non-epistemic values; anyone competent to redo a study will be too blinkered to consider alternatives. Furthermore, even if scientists *could* redo research with different values, we need an argument as to why they *ought* to do that; after all, they, too, have a claim that their autonomy be respected.

Similar worries arise with McKaughan and Elliott’s second suggestion: That audiences could try to assess whether changing values would make a difference to results. While we can picture how this might work in some cases, such reconstructions can require epistemic expertise; even if assessing the impact of different non-epistemic values on some finding simply requires tweaking some variable in a formula, most of us can’t tweak most variables in most formulae! So, attempts to reassess research will typically require expert input. Unfortunately, however, we now face a problem: It might be that certain sorts of values are simply “baked into” experts’ training and background assumptions; consider Lloyd’s discussion of the ubiquity of androcentric values in some research, mentioned in Section 16.1. When scientific communities are centered around some implicit value, it may be impossible to find “alternative experts” who can adopt the values of the audiences.

This leaves us with the third option: Simply to disregard research on the basis of its practitioners’ non-epistemic value commitments. I suggest, however, that this response might be problematic. While transparency might commit that, given the difference between our values and scientists’ values, we are not compelled to believe what they say, it doesn’t follow that we are therefore justified in disregarding what they say. After all, there is always a possibility that, even had they shared our values, they might have ended up with the same result. (Note that McKaughan and Elliott’s second response assumes this possibility.) So, in many cases, the proper response to revelations of value differences may be a kind of suspension of judgment. You might still think that this is a positive result of

transparency. I do not disagree. Still, it is important to stress that transparency can only take us so far. In particular, it is important to stress that learning the values implicit in research cannot necessarily justify us in *denying* that research. Even if interactive transparency allows us to identify values, it is not clear what we can or should do with that information.

Of course, there are answers to many of these challenges. In some cases, there are arguments that scientists' autonomy *ought* to be secondary to the audience's values (for example, Andrew Schroeder's arguments for why scientists should adopt "democratic values," in [Schroeder 2021](#)). Alternatively, we might think that, even if scientists cannot always "redo" their research, they can discuss with audiences the likely effects of changing various assumptions and commitments. Note, however, that all of these strategies require publics to trust scientists; for example, to take the results of public deliberation back into the lab or to be honest in assessing how different values would change their results. Calls for transparency are, however, often viewed as a response to failures of trust. What, then, is the relationship between trust, transparency, and non-epistemic values?

16.3 Trust or, or trust and, transparency?

It is common to read about a "crisis of trust" in science, or even expertise in general ([Eyal 2019](#)). These concerns are often linked to claims about the need for greater transparency in science. There is a seductive logic here, because we often think of mechanisms of transparency as an alternative to, or replacement for, trust ([O'Neill 2002](#), chap. 4). When we trust someone, we typically rely on them to act well without overt surveillance. It would seem strange for someone to say, "of course I trust my wife not to cheat, but I've placed a tracker app on her phone just to be sure." Typically, we demand transparency when and because we lack trust; in Thi Nguyen's pithy formulation, "transparency is surveillance." However, the links between trust and transparency are more complex than they seem. One set of concerns relates back to [Section 16.1](#): Sometimes, transparency can *destroy* well-grounded trust, as when false "folk philosophies of science" lead non-experts *away* from trusting climate scientists. In this section, I explore three further complexities: Around systematic mistrust; around notions of "enhanced trust"; and around the gap between what scientists do and what they say they do.

Consider the suspicious husband. We might say that he has replaced the need for trust with surveillance. However, this description misses something: The husband's beliefs about his wife's whereabouts depend on him believing that the app is accurate. However, this belief requires trusting the claims of marketers and the competence of app-developers. The app doesn't replace trust, it displaces it. A similar phenomenon is true in more complex scientific cases. When, for example, we require that the pharmaceutical industry pre-registers clinical trials, we don't simply replace trust with transparency; rather, we stop trusting the industry and, instead, start trusting the people and processes at the registry ([O'Neill 2002](#)). This phenomenon has an important upshot: Transparency mechanisms can only ever be a response to local or piecemeal lack of trust. They cannot be helpful responses to global or systemic concerns. If it is true that we are witnessing a crisis of trust in science (or expertise) in general, then making *parts* of the system transparent will not resolve the problem. (Note, then, there is an odd symmetry here: [Section 16.1](#) argued that, when the public *do* trust science, transparency can sometimes undermine well-grounded trust. Here, I have suggested that, when they do not trust science, transparency may fail to create trust.)

The second, related problem, concerns the distinction between "basic" and "enhanced" trust. We can distinguish two things we might mean when we say things like "I trust climate scientists": First, that we trust climate scientists to "do their job," in the sense of following the disciplinary norms of climate science; second, that we are willing to accept claims which have been established according to those norms ([John 2018](#)). These two aspects of epistemic trust can differ. For example, I might trust an astrologer to "do his job" (he really does calculate charts according to astrological norms) but still

refuse to trust his claims (because I think the norms of astrology are epistemically worthless). Clearly, both sorts of trust are required for full epistemic trust: I need to believe *both* that climate scientists play by the rules of climate science *and* that the rules of climate science are epistemically reliable. Unfortunately, while transparency might help bolster the first sort of trust, it does little for the second: Transparency gives us reason to believe that the rules were followed; however, if we already think the rules are bad, then we don't gain much by knowing that they have been followed.

Recent work by Gurol Irzik and Faik Kurtulmus on the relationship between epistemic trust and non-epistemic values in science has further complicated this picture, by distinguishing between "basic" and "enhanced" epistemic trust (Irizik and Kurtulmus 2019). Imagine a parent considering whether to vaccinate her child with the triple measles, mumps, and rubella (MMR) vaccine is concerned that standard ways of assessing the safety of vaccination are too tolerant of false positives and insignificantly sensitive to false negatives. This parent might concede that when some scientists say "the vaccine is safe," they have "done their job," in the sense that they have established this claim to the standards of virology. That is to say she has "basic trust." However, she may, nonetheless, refuse to defer to their testimony. She lacks "enhanced trust." In this case, her concern is not that the standards are straightforwardly epistemically unreliable (as in the case of the astrologer). Rather, it is that the standards (implicitly) reflect the "wrong" trade-off between different epistemic errors. To make the move to "enhanced trust" requires evidence that the non-epistemic values underlying the ways in which they handle epistemically underdetermined choices are aligned with the non-experts' values.

Irizik and Kurtulmus's account of trust raises many issues, but, for now, I will focus on the link to transparency. Often, when we encounter public hesitancy or resistance to scientific claims, we respond by making it transparent that we have followed standard procedures; for example, we might respond to hesitancy around COVID-19 vaccination by showing that, despite the urgency of the situation, we adopted standard testing protocols, rather than taking shortcuts (Kuter, Offit, and Poland 2021). This response will work for some audiences. However, it will not respond to concerns that the procedures themselves implicitly favor certain sorts of (problematic) non-epistemic values; for example, if you already think that the rules for testing vaccines are too tolerant of possible risks, you do not care that those protocols were followed. Transparency is not the best tool to tackle failures of enhanced epistemic trust.

The third complexity around trust and transparency concerns the relationship between "doing science" and "writing science" (Schickore 2008). Unlike religious gurus or charismatic leaders, scientists do not demand "blind trust." They do not say that some substance is carcinogenic or recommend that we ought to ban some substance and ask us to take their word for it. Rather, they offer data and arguments for their conclusions. An important feature of these justifications is that they are not really reports of *how* they came to their conclusions. Rather, they are closer to idealized reconstructions; for example, a scientist who says we ought to believe that the substance is carcinogenic on the basis of such-and-such an experiment may not report various failed experiments, or difficulties which arose when a lab worker took a new post, or she might present some hypothesis as arising from the literature, even if, in fact, inspiration struck her in the shower. Although annoying for historians of science, this is not a problem or fault in how we present scientific research. Rather, it is a function of the fact that a scientific paper is an argument for why the reader ought to believe a conclusion, not a report of how the author(s) came to that conclusion. (I promise this paper would be a lot worse if it narrated precisely how I came to my views, rather than what I now believe to be the best arguments for those views!)

Framed this way, demands for transparency can seem strange; they are a bit like playgoers demanding that they be allowed to see the lighting rig or the stage manager. Why, then, do we demand

to see backstage in science? There are three answers to this challenge. Two have already been canvassed above. First, while a good play can (and should) look very different from backstage, good science must be grounded in what scientists actually did; we need a guard against epistemic bad practice. Second, because the distinction between discovery and justification is porous, problematic features of the discovery process can feed into results and their justification, creating value-laden science (Elliott and McKaughan 2009). The third reason cuts across the other two: While scientists do justify their claims, these justifications typically don't provide non-experts with reasons or considerations they can readily understand (Goldman 2001). For most of us, scientists are practically equivalent to gurus, demanding our blind trust. However, even if we cannot assess the justifications scientists offer for their claims, we can, at least in principle, check whether the ways in which they arrived at those claims were in accordance with various sorts of processes which (we think) scientists ought to follow.

What transparency often does, then, is to allow us to ensure that processes are followed “properly” (John 2021, chap. 1). Section 16.1 already explored some problems with this checking: That there may be a gap between the processes we *think* scientists ought to follow and the processes they *actually* ought to follow. To conclude this section, it is worth noting a further problem: That if we think that scientists *ought* to defer to their audience's values, then they will often have good reason to go against standard processes and procedures. For example, in the case of the MMR vaccine above, the only way to respond to the worried parent's concerns might be to adopt different, non-standard testing procedures. In this sort of case, then, there may be a tension between what transparency measures are typically designed to track—whether we have followed the rules—and scientists acting in the most ethically responsive, and hence trustworthy, ways. Again, then, the relationships between transparency and trustworthiness may be very complicated.

16.4 Conclusions

The role and value of transparency in science is a vast topic. I have only scratched the surface. However, I do hope to have established three key claims. First, that the *very same* features of transparency which may make it attractive as a response to worries about “problematic values” in science—that it makes scientists act and reason as publics understand and appreciate—make it problematic as a response to worries about epistemic bad practice—because the public may just be confused on what counts as epistemic good practice. Second, that any account of what sorts of transparency are required to handle “problematic values” needs to avoid assuming that the influence of non-epistemic values is always a matter of conscious choice. Third, transparency may be a blunt tool for tackling mis- and dis-trust. These claims have important practical implications, because they suggest we should not always cheer calls for transparency. They also have important theoretical payoffs; clarifying why we value transparency has implications for how we understand trust and the role of values in justification. We might gain from being more transparent about the value of transparency.

Notes

- 1 For accounts of different forms of transparency in general, see Heald (2022); for a typology focused on the scientific case, see Elliott (2022).
- 2 The notion that transparency dissuades bad behavior has been a central argument for making political decision-making more open (Kogelmann 2021); as we will see, however, reading over from the political to scientific cases is complex.
- 3 Although these claims are as problematic in politics as in science, for a useful discussion of key themes around transparency and politics, see Chambers (2004).

- 4 For a helpful attempt at disambiguating various senses of the claim that science is “value-laden,” see Ward (2021).
- 5 These proposals often build on Helen Longino’s influential model of community-level deliberation (Longino 1990), albeit employing a very different account of how values enter science.

References

- Alexandrova, Anna. 2017. *A Philosophy for the Science of Well-Being*. Oxford University Press.
- Betz, Gregor. 2013. “In Defence of the Value Free Ideal.” *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Chambers, Simone. 2004. “Behind Closed Doors: Publicity, Secrecy, and the Quality of Deliberation.” *Journal of Political Philosophy* 12 (4): 389. <https://doi.org/10.1111/j.1467-9760.2004.00206.x>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2009. “How Do Disclosure Policies Fail? Let Us Count the Ways.” *The FASEB Journal* 23 (6): 1638–42. <https://doi.org/10.1096/fj.08-125963>.
- Douglas, Heather. 2000. “Inductive Risk and Values in Science.” *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Elliott, Kevin C. 2022. “A Taxonomy of Transparency in Science.” *Canadian Journal of Philosophy* 52 (3): 342–55. <https://doi.org/10.1017/can.2020.21>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. “How Values in Scientific Discovery and Pursuit Alter Theory Appraisal.” *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Eyal, Gil. 2019. *The Crisis of Expertise*. John Wiley & Sons.
- Goldman, Alvin I. 2001. “Experts: Which Ones Should You Trust?” *Philosophy and Phenomenological Research* 63 (1): 85–110. <https://doi.org/10.2307/3071090>.
- Heald, David. 2022. “The Uses and Abuses of Transparency.” In *This Obscure Thing Called Transparency: Politics and Aesthetics of a Contemporary Metaphor*, edited by Emmanuel Alloa. Leuven University Press. <https://doi.org/10.2307/j.ctv26dhjc9.5>.
- Holman, Bennett, and Sally Geislar. 2018. “Sex Drugs and Corporate Ventriloquism: How to Evaluate Science Policies Intended to Manage Industry-Funded Bias.” *Philosophy of Science* 85 (5): 869–81. <https://doi.org/10.1086/699713>.
- Hulme, Mike, and Jerome Ravetz. 2009, December 1. “‘Show Your Working’: What ‘ClimateGate’ Means.” *BBC News Online*. <http://news.bbc.co.uk/2/hi/8388485.stm>. Accessed 17 May 2024.
- Irzik, Gürol, and Faik Kurtulmus. 2019. “What Is Epistemic Public Trust in Science?” *The British Journal for the Philosophy of Science* 70 (4): 114566. <https://doi.org/10.1093/bjps/axy007>.
- John, Stephen. 2018. “Epistemic Trust and the Ethics of Science Communication: Against Transparency, Openness, Sincerity and Honesty.” *Social Epistemology* 32 (2): 75–87. <https://doi.org/10.1080/02691728.2017.1410864>.
- John, Stephen. 2021. *Objectivity in Science*. Cambridge University Press.
- Kogelmann, Brian. 2021. “Secrecy and Transparency in Political Philosophy.” *Philosophy Compass* 16 (4): e12733. <https://doi.org/10.1111/phc3.12733>.
- Koskinen, Inkeri. 2017. “Where Is the Epistemic Community? On Democratisation of Science and Social Accounts of Objectivity.” *Synthese* 194 (12): 4671–86. <https://doi.org/10.1007/s11229-016-1173-2>.
- Kovaka, Karen. 2021. “Climate Change Denial and Beliefs about Science.” *Synthese* 198 (3): 2355–74. <https://doi.org/10.1007/s11229-019-02210-z>.
- Kuter, Barbara J., Paul A. Offit, and Gregory A. Poland. 2021. “The Development of COVID-19 Vaccines in the United States: Why and How So Fast?” *Vaccine* 39 (18): 2491. <https://doi.org/10.1016/j.vaccine.2021.03.077>.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press.
- Lloyd, Elisabeth A. 2006. *The Case of the Female Orgasm: Bias in the Science of Evolution*. Harvard University Press. <https://doi.org/10.4159/9780674040304>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- McKaughan, Daniel J. and Kevin C. Elliott. 2013. “Backtracking and the Ethics of Framing: Lessons from Voles and Vasopressin.” *Accountability in Research* 20 (3): 206–26. <https://doi.org/10.1080/08989621.2013.788384>.
- Nguyen, C. Thi. 2022. “Transparency Is Surveillance.” *Philosophy and Phenomenological Research* 105 (2): 331–61. <https://doi.org/10.1111/phpr.12823>.
- O’Neill, Onora. 2002. *A Question of Trust: The BBC Reith Lectures 2002*. Cambridge University Press.

- Oreskes, Naomi, and Erik M. Conway. 2010. "Defeating the Merchants of Doubt." *Nature* 465 (7299): 686–7. <https://doi.org/10.1038/465686a>.
- Porter, Theodore. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.
- Schickore, Jutta. 2008. "Doing Science, Writing Science." *Philosophy of Science* 75 (3): 323–43. <https://doi.org/10.1086/592951>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Ward, Zina. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Willholt, Torsten. 2023. "Harmful Research and the Paradox of Credibility." *International Studies in the Philosophy of Science* 36 (3): 193–209. <https://doi.org/10.1080/02698595.2023.2273606>.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART III

Values, Science, and Democracy



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

SCIENTIFIC KNOWLEDGE AS A PUBLIC RESOURCE

Arguments and Challenges for a Democratic Approach to Values in Science

S. Andrew Schroeder

17.1 Introduction

The institutions of science ought to be guided by the needs, values, and preferences of the public—or, for short, science ought to be responsive to the public. This may seem like an uncontroversial idea, and indeed versions of it have been defended by many scholars and commentators of science. This includes a large and growing group of philosophers, sometimes described as proponents of “democratizing science,” who have argued that the (non-epistemic) values that shape scientific inquiry ought to in some way come from the public.¹ There are, however, many critical respects in which the current and historical institutions of science have not embraced this goal. Further, recent work by philosophers has shown how complex and underspecified this ideal is: It is not clear what it would mean for science to be appropriately responsive to the public or how scientists ought to pursue that goal.²

In this chapter, I will begin by reviewing the most straightforward argument for thinking that science ought to be responsive to the public, showing that the case for making science responsive to the public compares favorably to the case for making other critical institutions, such as a society’s educational or health care system, responsive to the public. I will then briefly describe several key respects in which the institutions of science have not adopted that goal. Finally, I will identify several unresolved challenges we face in understanding what it would mean for science to be appropriately responsive to the public.

17.2 Scientific knowledge as a public resource

There are many reasons one might offer in support of the idea that science ought to be responsive to the public, but the most straightforward argument is familiar—and, I think, convincing. Its core is the fact that scientific knowledge is one of the most important resources of any contemporary society. The shape of that resource—both what we know and what we don’t know—critically impacts nearly all aspects of our lives. Although the importance of science is universally acknowledged, it is worth briefly reminding ourselves of the many different kinds of impact it has.

The most obvious and direct impact of science on our lives is via the technology that it makes available to us. I am able to easily access vaccines that offer protection against polio and Covid-19, but not HIV or Chagas disease. I have a device that can quickly transmit images and sounds halfway around the world, but no ability to similarly transmit smells or tastes. We have consumer cars that

can, mostly reliably, navigate highways autonomously, but not the flying consumer cars that the science fiction of my childhood promised. Governments have devices that can quickly and precisely deliver explosives thousands of miles away, but no device that can do the same with vaccines.

More subtly, scientific knowledge impacts how efficiently or cost-effectively we can pursue our individual or social goals. The relative cost of different energy generation systems is not simply a brute fact about different technologies; it is something we can influence through research. The falling cost of solar energy generation, for example, is in large part due to research advances in that field. The cost-effectiveness of treating different illnesses depends in part on how much we know about those illnesses.

Scientific knowledge also improves our decision-making processes. Knowledge about the health consequences of smoking or the environmental consequences of driving allow me to make informed decisions about whether to engage in those activities. Governments use scientific knowledge to set food safety standards, building codes, and land use regulations. In other areas, though, a lack of scientific knowledge inhibits informed decision-making. Medical research has historically included mostly male research subjects, which has left us relatively ignorant of how certain health problems manifest in women (Johnson et al. 2014; see also Bluhm 2026 in this volume). Climate research has disproportionately studied the impact of climate change on wealthier countries, which has made it harder to make high confidence predictions concerning many regions in the Global South (Elabbar forthcoming; see also Lusk 2026 in this volume).

Those, I think, are the most obvious ways in which scientific knowledge shapes our lives, but its influence runs much deeper. In changing the possibilities that are available to us, scientific knowledge can change the moral status of our actions. The creation of an effective Covid-19 vaccine, for example, instantly created obligations (on the government to provide that vaccine, and on some private citizens to take it), while taking away certain moral entitlements (for certain employees to insist on remote working arrangements). Knowledge about increasing wildfire risk transforms the decision to live in the foothills around Los Angeles from mainly a personal choice to one that imposes burdens on society, who bear the costs of firefighting and rescue. Becoming aware of certain inequalities (such as the medical and climate research inequalities mentioned above) may create obligations to address or account for the resultant disadvantages.

Finally, scientific knowledge can, in a very real way, change how we think and feel about the world, or even what thoughts, attitudes, and identities we are able to have. Consider evolution by natural selection, the discovery of the atom, or Einsteinian relativity. These discoveries radically changed the way many people conceived of the world and humanity's place in it. Research on race and gender has changed the way many people view those categories, opening up new possibilities for how societies might be organized. And evolving views in medicine and psychology have created new possibilities for individuals—new social and personal identities they are now able to adopt (Hacking 2006).

That, then, is an incomplete list of the ways that the body of scientific knowledge—both what it includes and what it excludes—shapes our lives: It impacts our material circumstances, changes the cost-effectiveness of different endeavors, can facilitate better and more informed decision-making, changes the moral status of our actions, and shapes what thoughts, attitudes, and identities we are able to have. It is hard to think of many other institutions that have a similarly pervasive impact. Candidates might include a society's basic political or economic structure, or its educational or health care systems. It is widely accepted that those systems ought to be responsive to the needs, values, and preferences of the public, and so the case for making science responsive to the public seems strong.

Two further features of science make the case even stronger. First, most citizens are not in a position to meaningfully shape our society's body of scientific knowledge. If you find that our collective knowledge of some disease, exoplanet, or chemical process falls short of what you want it to be, there

is probably little that you can do to change the situation. Most people lack the expertise to directly conduct such research, and in any case scientific research is often expensive and time-consuming.

Second, scientific knowledge has the features that characterize what economists call a public good. Public goods in the economic sense are non-rivalrous, meaning that one person's use of the good does not diminish others' ability to also use it; and they are non-excludable, meaning that it is hard to prevent specific people from making use of the good. Unpolluted air is an example: The fact that I breathe unpolluted air doesn't prevent others in my community from also breathing unpolluted air, and there is no way to stop other community members from enjoying that benefit. Scientific knowledge is often cited as a paradigm example of a public good ([Dalrymple 2003](#); [Hess and Ostrom 2006](#); [Stiglitz 1999](#)). My studying up on quantum mechanics does not prevent others from also studying up on quantum mechanics. And once scientific knowledge begins to be disseminated, it is very hard to stop specific people or groups of people (for example, those who have not paid a fee) from accessing that knowledge. Economists generally agree that private markets tend not to effectively supply public goods ([Anomaly 2015](#)). Thus, in order to produce an adequate supply of public goods, some government intervention is required. We indeed see this in the case of science. Governments have for a long time been one of the primary funders of scientific research, and governments establish intellectual property regimes, which (among other things) artificially create a kind of excludability by forbidding people from using certain kinds of information.

Putting these three ideas together: Scientific knowledge is a resource that has a pervasive impact on our lives; most individuals have little ability to meaningfully change the shape of that resource; and, in light of its character as a public good, governments already exert a significant influence on the development of science. Those three facts ground an extremely strong reason to insist that science be responsive to the public.³

17.3 Is science responsive to the public?

The previous section argued that we have very strong reasons to want science—in particular our body of scientific knowledge—to be shaped in a way that is responsive to the needs, values, and preferences of the public. But what does that call for? What would it mean to shape scientific knowledge in that way? There are two key kinds of decisions that directly and predictably determine the shape of that resource. First and most obviously are decisions about which topics get studied, or which questions get investigated. If scientists choose to study biological phenomena rather than astronomical phenomena, it is likely that we will come to know more about biology than astronomy. When scientists focus their attention on treatments for cancer over treatments for dengue, it is likely we will develop treatments for cancer sooner than treatments for dengue. Second, and more subtly, are value-laden methodological decisions about *how* scientists investigate particular topics or questions. When scientists define terms, manage inductive risk, choose classification systems, and build models (among other things), their decisions can reflect values; they can (intentionally or not) serve the interests of some over the interests of others. When a climate modeling team chooses which kinds of accuracy to prioritize, an economist chooses how to quantify inequality, or an epidemiologist chooses what disease classification system to employ, these decisions can make the resulting research more or less useful for particular people or for particular purposes.⁴ As an initial attempt, we might therefore propose that making our body of scientific knowledge responsive to the public means ensuring that decisions about what to research and value-laden methodological decisions about how to research those topics are made in ways that are guided by the public's needs, values, and preferences.⁵

Before discussing why things are much more complicated than this initial gloss suggests, it is worth emphasizing that science, as an institution, is currently very far from achieving that sort of alignment, or even accepting it as a goal. A large (and increasing) share of scientific research funding

comes from industry sources.⁶ Unsurprisingly, we have clear evidence that industry-funded research tends to be shaped in ways that increase profits, rather than in ways that reflect the public's values or respond to the public's needs.⁷ Of course, a great deal of scientific research is funded by governments. Since elected representatives and appointed officials are supposed to be accountable to the public, we might expect public funding to be allocated in a way that is responsive to the public. However, at least in the US, much public funding for science has been allocated under a model that delegates key funding decisions to scientists, and grants for so-called basic research have generally placed little emphasis on alignment with the public's needs, values, or preferences (Douglas 2009; Douglas and Branch 2024).⁸

Turning to the value-laden methodological decisions that shape research, in most cases codes of professional ethics grant scientists significant latitude in making them (Schroeder 2022a); they therefore do not instruct scientists to make those decisions with an eye to what the public needs or values. A great deal of literature, largely owing to work by feminist scholars, has shown that when we allow scientists latitude in making these decisions, they frequently do so in ways that systematically ignore the interests of large segments of the population, and therefore can't reasonably claim to be responsive to the public as a whole (Kourany 2010; see also Intemann 2026 in this volume).

It therefore is clear that, currently, neither research funding decisions nor value-laden methodological decisions are generally made in ways that are guided by the public's needs, values, or preferences. Why isn't the right response to this to simply direct decision-makers to focus on the public when making these decisions? I will briefly discuss three unresolved philosophical debates, which show why the solution is not so simple.

17.3.1 Challenge 1: Thinking systemically

If science should be guided by the public's needs, values, or preferences, it might seem as if key actors should have that goal in mind when making decisions that shape science—that a funding agency should aim to support research projects that promise to have the greatest expected impact on the public good, and that individual scientists should aim to make value-laden methodological decisions in a way that reflects what the public values. However, a bit of reflection shows that that isn't necessarily correct. There are many contexts in which the best way for a system to achieve some goal is for key actors in that system to pursue different and apparently conflicting goals. Consider, for example, a criminal trial. The goal of a trial is, roughly, to convict the defendant if and only if they are guilty. Yet to achieve that goal, we assign key players different goals which appear to be in tension with that overarching goal: Defense attorneys are to (within certain limits) make the case that their client is innocent, and prosecutors are to (within more extensive limits) make the case that the defendant is guilty. We do this because we believe that the overall system will more reliably achieve our goal than would an alternative system which directed the attorneys on both sides to directly pursue the goal of convicting the defendant if and only if they are guilty.

Recent work in political theory has argued that a similar point applies to democratic systems. Proponents of deliberative democracy put forward an ideal in which people collectively engage in rational deliberation on terms of equal respect and inclusion to make the decisions that will govern their community (Bächtiger et al. 2018). Early proponents of deliberative democracy, accordingly, sought to design fora in which those values were embodied to the greatest extent possible. They created "mini-publics," where a representative subset of the public was brought together; put in a structured environment designed to promote inclusion, respect, and equal consideration; and then encouraged to reach decisions in ways that emphasized deliberation and reason-giving (over, say, coercion and manipulation).⁹

Recent scholarship, however, has emphasized that if our goal is to produce a system that embodies the values of equality, respect, inclusion, and rational deliberation, it may be best to have subparts of that system that intentionally fail to embody those values (Mansbridge et al. 2012). For example, if a cultural group consistently has its voice drowned out in public debate, it might make sense to create venues reserved exclusively for that group. Considered in isolation such venues would, obviously, fail to embody the values of inclusion and equality, but they might contribute to inclusion and equality at the level of the overall system by promoting voices that otherwise would not be heard. Similarly, disruptive activism often aims to achieve its ends through power and coercion, rather than rational persuasion. Nevertheless, some instances of disruptive activism may contribute to inclusive rational deliberation at the system level, by ensuring that certain viewpoints which would otherwise have been ignored are placed on the agenda for consideration.

A similar evolution is happening in the literature on value-laden methodological decisions in science. Earlier proponents of a “democratic values” approach tended to endorse making individual methodological decisions in a way that was guided by the public’s needs, values, or preferences.¹⁰ Several recent authors have pushed back on this. They have endorsed the broader goal of making science responsive to the public but proposed indirect paths to achieving that goal. Pamuk (2021) proposes an adversarial “science court,” and Hilligardt (2023) emphasizes the importance of partisan science—science conducted from a particular ideological perspective.

Future work will need to determine how to best integrate the earlier more direct approach with the later indirect approach. In thinking about democracy more generally, there are some contexts where it is important to insist on deliberations that approximate the ideal as far as possible: We generally want jury deliberations to be governed by norms of inclusion, respect, equality, and rational deliberation. There are other contexts where it is appropriate to support temporary departures from the ideal to compensate for deficiencies elsewhere in the system—for example, creating exclusive fora for marginalized voices. And, finally, there are contexts where permanent departures are potentially justified. Certain adversarial proceedings may be good examples. It seems reasonable to think similar conclusions will apply in the case of science: There may be certain kinds of decisions where scientists should aim to be guided by the public’s values; other cases where we should encourage certain actors to depart from that ideal temporarily, to compensate for flaws elsewhere in the system; and some cases where we should construct sub-systems that consistently reject the goal of being responsive to the public.

17.3.2 Challenge 2: Responding to a diverse public

To this point in the chapter, I have been talking as if there is a unified set of public needs, values, and preferences. But, of course, that is rarely the case. The public is composed of individuals who have different needs, values, and preferences. How should we proceed when members of the public want or need different things out of science?¹¹ Suppose, for example, that the public is divided about energy policy. One group prefers prioritizing research into improving the efficiency of solar power generation, while a second prefers working to make nuclear power safer. How should we allocate research funding in the face of such disagreement, if our goal is to be directly responsive to the public?

In a democratic society, we typically manage disagreements like this either by deferring to the majority or by choosing some compromise that reflects public support for each option.¹² If, for example, two-thirds of the public prefers the solar option, then we might allocate two-thirds of the funding to that project, and one-third to the nuclear project. Let us assume that such an approach is a legitimate way of managing disagreement concerning many ordinary policy decisions—e.g., setting tax rates, vehicle speed limits, and so forth. As Thoma (2024) points out, scientific research raises special concerns. When we are deciding what sort of science to support, the “losers” don’t simply fail to get the

policy outcome they want (e.g., more funding for research into nuclear technology); they also suffer a downstream *epistemic* inequality. Because less research will have been done on nuclear technology, proponents of nuclear technology will have less information at their disposal to use in future policy debates. That epistemic inequality seems troubling from a democratic perspective. Although democracy does require us to accept that we will not always get our way, fundamental to the democratic process is that each of us gets a fair chance to make the case for our preferred outcome. And, at least in many situations, making a compelling case for some option requires knowledge about that option.¹³

Outcome-based democratic norms, then, tell us to defer to the majority on certain key decisions. But epistemic norms central to democracy direct us to preserve the minority's ability to advocate for their position in the future. It is not clear now to navigate these apparently conflicting requirements. Thoma focuses on the case of social scientific indicators, such as wellbeing or cost-of-living indices. She argues that while we might make policy decisions grounded in indices that reflect majority values, scientists should publish “dashboards” of unaggregated sub-indicators, or a range of different composite indices. Such an approach would give the minority the information they need for future advocacy.

Thoma's proposal gives rise to a number of concerns, as she notes. But perhaps the most serious is that it may not travel well to other scientific fields. With social scientific indicators, publishing a plurality of composite indices may be as simple as aggregating a common data set in different ways—by, for example, weighting or standardizing various components differently. In many other scientific contexts, though, the kind of pluralism Thoma proposes can be much more costly. Re-running a complex climate model using different parameters can take weeks or months of supercomputer time. Measuring vaccine effectiveness in a different way can require recruiting a different cohort of research subjects and conducting an entirely new study (Schroeder 2025). In such cases, we face a trade-off: In order to achieve the kind of epistemic equality democracy ideally calls for, we would have to make significant sacrifices to scientific progress on other questions.

17.3.3 Challenge 3: Diversity within the extended scientific community

The previous section focused on a challenge that arises due to diversity among the public. Another important set of challenges arise from diversity within the institutions of science. Scientists conduct research on everything from sub-atomic particles to dinosaurs to economic prosperity. They work at public universities, private colleges, corporations, museums, government agencies, NGOs, private foundations, and via community partnerships. The extended scientific community includes scientific journals, large corporations, small start-ups, granting agencies, science journalists, science educators, museum curators, professional associations, regulatory bodies, policy advisors, hobbyists, and lobbyists—just to name a few.

When thinking about the extent to which science ought to be responsive to the public, it seems obvious that this diversity must matter. Consider a scientist working at a government agency conducting research relevant to an upcoming legislative hearing. It seems quite natural to say that she ought to be guided by the public's values. But what about a scientist employed by an environmental advocacy organization? Or a retiree conducting self-funded research in their back yard? Or an astrophysicist at a religiously affiliated university doing research with no foreseeable practical relevance? Perhaps these researchers also ought to be guided by the public's values, but the case doesn't seem nearly as strong. In some of the examples, this is because the positive case for aligning science with the public's values isn't as weighty. (If the astrophysicist's research isn't foreseeably decision-relevant for the public, then is it really important that the public have a voice in shaping it?) In others, it is because there are strong countervailing reasons. (Doesn't the retiree have a right to spend their time and money as they wish? Don't we have good reason to grant religious institutions significant independence from democratic control?)¹⁴ And in still other examples, it is due to the kinds of systemic factors discussed

above. (Perhaps the environmental advocacy organization serves as a counterbalance to lobbying from fossil fuel companies.)

Further, as that last example illustrates, taking a systemic perspective means that even if we have a clear overall goal or ideal for science, we can't in a straightforward way assign that goal or ideal to individual scientific actors. Whether a scientist ought to make methodological decisions in a way responsive to the public, or whether a scientific journal ought to insist on that as a condition of publication, or whether grant reviewers ought to look for that in funding applications, or whether science journalists ought to highlight aspects of research that meet (or fail to meet) that standard—what each of those actors ought to do may depend on the role they play within the larger system. To date, the philosophical literature has not effectively engaged with the diversity within the institutions of science; it has either put forward broad goals for science, or else discussed narrow slices of the scientific community. Future work could benefit from taking a more comprehensive approach—looking at the many different actors and institutions that comprise the scientific community, and considering how they can collectively contribute to the goal of making science, as a whole, responsive to the public.

17.4 Conclusion

As the previous sections show, there is a great deal more work to be done—work which will draw heavily on the philosophy of science and political philosophy—to understand what it would mean for science to be responsive to the public. Nevertheless, even if we are far from fully specifying that ideal, there is plenty that we can do right now to move in the right direction; to make science *more* responsive to the public than it currently is. We have examples, such as research into Covid-19, that show how science can quickly and effectively orient itself to address urgent public needs ([Callender 2020](#)). Though we may not yet know what an ideal funding model would look like, the increasing share of research funding that comes from industry suggests that we ought to either increase other sources of funding for science, or else find ways to reorient industry-funded science toward the public good. ([Holman and Elliott 2018](#) review a number of suggestions.) As we saw above, there are reasons to think that certain kinds of researchers should be free to adopt an explicitly partisan or ideological approach. But scientists who *want* to make value-laden methodological decisions in ways that are guided by the public's values should be encouraged to do so. They need not wait for philosophers to identify the optimal way to identify the appropriate public needs, values, and/or preferences. Directly consulting the public or inviting them into the research process as collaborators would surely constitute a significant improvement on the status quo ([Alexandrova and Fabian 2022](#)), in addition to securing a number of other important benefits ([Schroeder 2022b](#)).

Acknowledgments

For comments on earlier versions of this chapter, I thank Greg Lusk and Kevin Elliott. I also benefited from discussion at the 2023 Workshop on Intervening in Nature, sponsored by UC San Diego's Institute for Practical Ethics.

Notes

- 1 See, for example, [Douglas \(2005\)](#), [Kitcher \(2011\)](#), [Intemann \(2015\)](#), [Schroeder \(2021, 2022a\)](#), [Lusk \(2021\)](#), [Pamuk \(2021\)](#), and [Alexandrova and Fabian \(2022\)](#).
- 2 One important question which I elided above concerns whether science should be responsive to the public's values, needs, preferences, or something else; and whether it should be responsive to the public's actual attitudes or to some idealized version of their attitudes (e.g., to what the public *would* value if it were well-informed). The existing literature has not reached any consensus on this question. For reasons of space, I will unfortunately have to set it aside in this chapter. For the same reason, I will also largely set aside questions

concerning how these ideas could be put into practice. Karen Kovaka's (2026) contribution to this volume shows a few of the (considerable) challenges involved in implementation.

- 3 Of course, even very strong reasons can sometimes be outweighed by still stronger ones. Below, I will briefly flag some considerations that could potentially qualify.
- 4 For general overviews, see Elliott (2022) and Biddle and Kukla (2017). Many of the chapters in this volume offer further discussion.
- 5 This is, roughly, what many philosophical proponents of “democratizing science” propose. See, for example, Kitcher (2011), Lusk (2021), Intemann (2015), Douglas (2005), and Schroeder (2021). It is worth noting that the literature tends to handle questions about research funding/priorities separately from questions about value-laden methodological decisions. Perhaps surprisingly, philosophers of science have written much more about the latter topic than about the former. (Non-philosophers have largely focused on the former topic.)
- 6 For data about the US, see Pece and Anderson (2024).
- 7 See, for example, Oreskes and Conway (2010) on research funded by tobacco and fossil fuel companies. Lexchin et al. (2003) show that industry-funded toxicological research is systematically more likely to produce results favorable for its sponsors. And note the persistent complaints directed to pharmaceutical companies, for prioritizing profitable drugs over those with the greatest health impact (Ramsay 2001). See also Bravo (2026) in this volume.
- 8 Many scientists argued that this arrangement—whereby scientists make funding decisions without considering the downstream practical impact of their research—would ultimately yield immense public benefits (and thus could reasonably claim to be indirectly guided by the public's needs, values, and preferences). But, as Douglas and Branch note, there is no empirical research to support the claim that such a model is as effective at meeting the public's needs as alternative funding models which do explicitly consider downstream practical impact. See also Callender (2020).
- 9 For one version of this, see Fishkin (2018).
- 10 See, e.g., Kitcher (2011), Lusk (2021), and Schroeder (2021).
- 11 The plurality of views held by the public gives rise to a number of problems, beyond those I can discuss in this chapter. See Kovaka (2026) and de Melo-Martín (2026) in this volume for examples.
- 12 Le Bihan (2024) worries that traditional majoritarian solutions can give rise to problems of marginalization. I respond to this worry in an unpublished work-in-progress.
- 13 See also Hillgardt and Wilholt (2026, §§3–4), Kurtulmus and Irzik (2017), and Irzik and Kurtulmus (2024).
- 14 These are some of the considerations, referenced in note 3, above, which could potentially outweigh the reasons in favor of making science responsive to the public.

References

- Alexandrova, Anna, and Mark Fabian. 2022. “Democratizing Measurement: Or Why Thick Concepts Call for Co-production.” *European Journal for Philosophy of Science* 12: 7. <https://doi.org/10.1007/s13194-021-00437-7>.
- Anomaly, Jonathan. 2015. “Public Goods and Government Action.” *Politics, Philosophy & Economics* 14 (2): 109–28. <https://doi.org/10.1177/1470594X13505414>.
- Bächtiger, André, John S. Dryzek, Jane Mansbridge, and Mark Warren. 2018. “Deliberative Democracy: An Introduction.” In *The Oxford Handbook of Deliberative Democracy*, edited by Bächtiger, André, John S. Dryzek, Jane Mansbridge, and Mark Warren. Oxford University Press.
- Biddle, Justin B., and Rebecca Kukla. 2017. “The Geography of Epistemic Risk.” In *Exploring Inductive Risk*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Bluhm, Robyn. 2026. “‘Every Cell Has a Sex’: Sex Essentialism in Biomedical Research.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Bravo, Pedro. 2026. “Values and Industry-Funded Research.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Callender, Craig. 2020. “COVID-19 Is Making Science Better.” *Issues in Science and Technology*, June 11. <https://issues.org/covid-19-is-making-science-better/>.
- Dalrymple, Dana. 2003. “Scientific Knowledge as a Global Public Good: Contributions to Innovation and the Economy.” In *The Role of Scientific and Technical Data and Information in the Public Domain*, edited by Julie M. Esanu and Paul F. Uhrlir. The National Academies Press.
- de Melo-Martín, Inmaculada. 2026. “FDA Evidentiary Standards and the Need to Attend to Stakeholders' Values.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.

- Douglas, Heather. 2005. "Inserting the Public into Science." In *In Democratization of Expertise? Exploring Novel Forms of Scientific Advice in Political Decision-Making*, edited by Sabine Maasen and Peter Weingart. Springer.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather, and T. Y. Branch. 2024. "The Social Contract for Science and The Value-Free Ideal." *Synthese* 203: 40. <https://doi.org/10.1007/s11229-023-04477-9>.
- Elabbar, Ahmad. forthcoming. "Varying Evidential Standards as a Matter of Justice." *British Journal for the Philosophy of Science*. <https://doi.org/10.1086/727429>.
- Elliott, Kevin C. 2022. *Values in Science*. Cambridge University Press.
- Fishkin, James S. 2018. *Democracy When the People Are Thinking: Revitalizing Our Politics through Public Deliberation*. Oxford University Press.
- Hacking, Ian. 2006. "Making Up People." *London Review of Books* 28 (16), August 17. <https://www.lrb.co.uk/the-paper/v28/n16/ian-hacking/making-up-people>.
- Hess, Charlotte, and Elinor Ostrom. 2006. "Introduction: An Overview of the Knowledge Commons." In *Understanding Knowledge as a Commons: From Theory to Practice*, edited by Charlotte Hess and Elinor Ostrom. MIT Press.
- Hillgardt, Hannah. 2023. "Partisan Science and the Democratic Legitimacy Ideal." *Synthese* 202: 135. <https://doi.org/10.1007/s11229-023-04370-5>.
- Hillgardt, Hannah, and Torsten Uhlholt. 2026. "Science and Democracy." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott, and Richards Ted. Routledge.
- Holman, Bennett, and Kevin C. Elliott. 2018. "The Promise and Perils of Industry-Funded Science." *Philosophy Compass* 13 (11): e12544. <https://doi.org/10.1111/phc3.12544>.
- Intemann, Kristen. 2015. "Distinguishing Between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Intemann, Kristen. 2026. "Feminist Challenges to the Value-Free Ideal of Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Irzik, Gürol, and Faik Kurtulmus. 2024. "Distributive Epistemic Justice in Science." *British Journal for the Philosophy of Science* 75 (2): 325–45. <https://doi.org/10.1086/715351>.
- Johnson, Paula A., Therese Fitzgerald, Alina Salganicoff, Susan F. Wood, and Jill M. Goldstein. 2014. "Sex-Specific Medical Research: Why Women's Health Can't Wait. A Report of the Mary Horrigan Connors Center for Women's Health & Gender Biology at Brigham and Women's Hospital." <https://www.brighamandwomens.org/assets/bwh/womens-health/pdfs/connorsreportfinal.pdf>.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Kourany, Janet. 2010. *Philosophy of Science After Feminism*. Oxford University Press.
- Kovaka, Karen. 2026. "Measuring the Public's Values: Are the Data Adequate for Purpose?" In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Kurtulmus, Faik, and Gürol Irzik. 2017. "Justice in the Distribution of Knowledge." *Episteme* 14 (2): 129–46. <https://doi.org/10.1017/epi.2015.65>.
- Le Bihan, Soazig. 2024. "How to Not Secure Public Trust in Science: Representative Values Versus Polarization and Marginalization." *Philosophy of Science* 91 (5): 1128–38. <https://doi.org/10.1017/psa.2023.119>.
- Lexchin, Joel, Lisa A. Bero, Benjamin Djulbegovic, and Otavio Clark. 2003. "Pharmaceutical Industry Sponsorship and Research Outcome and Quality: Systematic Review." *BMJ* 326: 1167–70. <https://doi.org/10.1136/bmj.326.7400.1167>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Lusk, Greg. 2026. "Philosophical Approaches to Values in Climate Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Mansbridge, Jane, James Bohman, Simone Chambers, et al. 2012. "A Systemic Approach to Deliberative Democracy." In *Deliberative Systems*, edited by John Parkinson and Jane Mansbridge. Cambridge University Press.
- Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Pamuk, Zeynep. 2021. *Politics and Expertise: How to Use Science in a Democratic Society*. Princeton University Press.
- Pece, Christopher V., and Gary W. Anderson. 2024. *Analysis of Federal Funding for Research and Development in 2022: Basic Research*. NSF 24-332. National Center for Science and Engineering Statistics, U.S. National Science Foundation. <https://ncses.nsf.gov/pubs/nsf24332>.

- Ramsay, Sarah. 2001. "No Closure in Sight for the 10/90 Health-Research Gap." *The Lancet* 358 (9290): 1348. [http://doi.org/10.1016/s0140-6736\(01\)06465-0](http://doi.org/10.1016/s0140-6736(01)06465-0).
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Schroeder, S. Andrew. 2022a. "An Ethical Framework for Presenting Scientific Results to Policy-Makers." *Kennedy Institute of Ethics Journal* 32 (1): 33–67. <https://doi.org/10.1353/ken.2022.0002>.
- Schroeder, S. Andrew. 2022b. "Diversifying Science: Comparing the Benefits of Citizen Science with the Benefits of Bringing More Women into Science." *Synthese* 200: 306. <https://doi.org/10.1007/s11229-022-03774-z>.
- Schroeder, S. Andrew. 2025. "The Influence of Values on Medical Research." In *Oxford Handbook of Philosophy of Medicine*, edited by Alex Broadbent. Oxford University Press.
- Stiglitz, Joseph E. 1999. "Knowledge as a Global Public Good." In *Global Public Goods*, edited by Inge Kaul, Isabelle Grunberg, and Marc A. Stern. Oxford University Press.
- Thoma, Johanna. 2024. "Social Science, Policy and Democracy." *Philosophy and Public Affairs* 52 (1): 5–41. <https://doi.org/10.1111/papa.12250>.

MEASURING THE PUBLIC'S VALUES

Are the Data Adequate for Purpose?

Karen Kovaka

18.1 Introduction

If we want to know how different diseases or environmental hazards impact human health, we have to be able to measure the health losses these diseases and hazards cause. And measuring health losses, in turn, requires researchers to make a series of difficult normative decisions. Here are just two of them: Should they use temporal discounting, and count years of life lost in the future as less valuable than those lost in the present? Should they use age-weighting and treat certain years of life as more valuable than others? Different ways of answering these questions lead to different estimates of health losses. Indeed, one influential study reported that mental health conditions made up over 10% of the global burden of disease ([Murray and Lopez 1996](#)). But the same study made clear that this estimate depended on choices to age-weight and temporally discount. Without age-weighting and discounting, the estimate dropped to 5.6% ([Murray and Lopez 1996](#); [Schroeder 2021](#)).

This example illustrates a general fact about science: Across disciplines, there are times when conducting research and reaching conclusions requires scientists to make *unforced choices*, that is, decisions that are not fully determined by epistemic considerations, which they could reasonably make otherwise, and which might affect the conclusions they reach.¹ A partial list of such unforced choices includes decisions about which specific research questions to investigate, where to set evidential thresholds, how to operationalize thick concepts, what to represent in model development, and how to interpret model results. In many cases, philosophers have argued, there is no way to make these choices without appealing to social and political values. These insights have motivated a growing focus in philosophy of science on what has been called the “new demarcation problem” ([Holman and Wilholt 2022](#); [Resnik and Elliott 2023](#)), which is the problem of distinguishing the appropriate from inappropriate influence of values on science.

One influential response to the new demarcation problem holds that it is appropriate for social and political values to influence science if that influence is democratic. According to this *democratization view*, when scientists make unforced choices on the basis of social and political values, those values should be the ones held by the public, rather than scientists’ own personal values.² Why? Because when the content of scientific conclusions depends on unforced choices, a democratic foundation is supposed to confer a kind of legitimacy or trustworthiness on those conclusions that would otherwise be absent. Despite varying in many other ways, all versions of the democratization view share this fundamental justification.

The growing philosophical literature on the democratization view is focused on developing and evaluating its normative claims, including the idea that democratization lends legitimacy to science. But there is an empirical aspect of the view that has been comparatively neglected. How are scientists supposed to know what the public's values are, in the first place? The democratization view answers this question by appealing to social science. After all, social scientists have a suite of methods that are capable of generating data about what people value. So, defenders of the view say, scientists and scientific institutions should work with social scientists to figure out what the public's values are. No doubt the actual process of measuring values will be complex and difficult, but the democratization view assumes that it can be done, and that the resulting data will be adequate for the purpose of legitimizing scientific conclusions.

It is precisely this claim that I want to examine more carefully. Are social scientific data adequate for the purpose of making scientific conclusions legitimate in the way the democratization view claims?³ In order to answer this question, we will need a clearer understanding both of the relevant social scientific data, and of the legitimizing purpose for which the democratization view wants to use these data. From this discussion, I will show that there are limitations of the data that make them inadequate for conferring legitimacy in the sense that defenders of the democratization view hope for. This conclusion does not mean we should reject the democratization view, but it does mean that we should refine it in light of the nature of the available data.

18.2 Characterizing the data

At present there is not very much research that does what the democratization view calls for—identifying the public values that are relevant to the actual unforced choices that scientists make in the course of doing their work. There are some exceptions to this generalization (e.g., [Alexandrova and Fabian 2022](#)), but most of the data scientists would need in order to make decisions on the basis of public values do not exist. Fortunately, social scientific methods for identifying values are highly developed, so we can still characterize the data that these methods produce. To do this, I will start with a brief discussion about values, which is what the data are supposed to measure or capture. Then I will describe the methods researchers use to measure values and the nature of the data these methods produce.

The understanding of values at play in the democratization view is quite broad. Kevin Elliott has characterized values as things that are “valuable or worthy of pursuit” (2017, 11). This definition can be further narrowed to values that are non-epistemic, i.e. social or political, what Rachel Baker and collaborators refer to as “higher level principles, moral or socio-cultural standards that people hold as good or right” (Elliott 2017, 2). Typical examples of values so defined include protecting human autonomy, alleviating suffering, increasing standards of living, and preserving biodiversity. When we talk about measuring the public's values, the relevant sense of “value” is closer to beliefs or attitudes about these principles, standards, and desirable things, such as attitudes about their relative importance (see Elliott and Korf 2024).

Values, or attitudes toward values, come into the scientific research process when unforced choices, or *value judgments*, are made because they are justified by certain values as opposed to others (Ward 2021). Compared to the value of preserving biodiversity, for example, a judgment that is justified by commitment to this value might be an ecologist's choice to operationalize ecosystem health in terms of the number of indicator species an ecosystem contains, rather than in terms of other metrics, such as primary productivity or the rate of nutrient cycling. This distinction between values and value judgments is important because while value judgments are what scientists *make*, it is the strength of the public's commitment to different values that social scientific methods are well-equipped to *measure*.

The sheer number of scientific value judgments mean that it is not possible, let alone desirable, to ascertain public opinion about each and every one of them (Elliott 2022, 40–41). Not only that, another important difference between having values and making value judgments is that the latter, but not the former, often requires scientific expertise that members of the public do not possess (Douglas 2009, 73). So, even if scientists have detailed information about the public's values, this does not necessarily mean it will be easy or straightforward to make judgments about age-weighting, temporal discounting, measures of ecosystem health, etc. Instead, the democratization view must accept that there will be gaps between the public's values and scientists' value judgments, gaps which can only be bridged via normative reasoning and argument. Sometimes the process of bridging values gaps may be fairly straightforward or obvious, but in other cases it will be complex and contestable.

One potential way of making these values gaps smaller and more tractable is to measure the public's values in ways that are contextually specific. That is, rather than measuring how important the public thinks preserving biodiversity is in general, social scientists could measure how important this value is to the public in particular domains. Indeed, asking about values while *fully* specifying the relevant decision context is equivalent to asking what value judgment a subject thinks should be made. As noted above, this suggestion is impractical, but there still may be some amount of context that is useful and important to include when measuring public values. Such contextually specific values data would partially reduce the need for scientists to make educated guesses about how to apply more general values data to their individual research projects. At the same time, contextually specific measures come with drawbacks and downsides of their own. As is so often the case with scientific measurement, no single metric gives us everything we want. Tradeoffs and compromises are the order of the day. To see this, let's move to a survey of the main types of social scientific methods for collecting data about the public's values.

There is a wealth of different social scientific methods that can provide this kind of data, but for our purposes, the diversity of methods falls into five main groups.

- 1 *Preference elicitation*. This is an umbrella term for a set of methods designed to quantify how much people value different outcomes. Subjects might be asked to arrange different health states along a scale of best to worst, or they might be asked how many years of full health they think would be worth x years in a diminished health state. While preference elicitation is most highly developed in the field of health economics, the methods themselves can be used more broadly. Daniel Hausman (2023) provides a more comprehensive review of these methods.
- 2 *Surveys*. Surveys are used in many fields, including environmental and health economics, bioethics, social psychology, and sociology, to identify and measure values. Researchers develop measurement instruments known as scales, in which values are treated as latent variables. Subjects are given a series of statements, to which they indicate their level of agreement on a scale, from 1 to 5, 1 to 7, etc. Different subsets of these statements correspond to different values, and the strength of subjects' commitment to different values is inferred from the answers they give to the different survey items (see, e.g., Evans and Schairer 2024; Stange et al. 2022).
- 3 *Interviews*. Compared to surveys and preference elicitation, interviews provide more open-ended, qualitative data and are more time-consuming to conduct. They can be more or less structured, depending on the research aims. Interviews are also frequently combined with the other methods described here, often as a way of collecting preliminary data, or to round out the sort of data collected by any given research project.
- 4 *Deliberative methods*. This is another umbrella term for methods that involve (a) providing participants with more detailed information about decision scenarios, and (b) opportunity for discussion among participants. These methods are often used when policymakers seek input from the public about specific policy choices. In such cases, the argument goes, public input is not particularly

useful unless it is well-informed, so deliberative methods produce better results than surveys or preference elicitation. Deliberative Polling, as developed by Stanley Fishkin (2009), is a paradigmatic example of this sort of method.

- 5 *Community-based participatory research (CBPR)*. CBPR (and other related methods) adds an iterative element to the process of measuring values. Researchers and broader communities build relationships, allowing researchers to pursue research questions that fit well with community needs. The researcher-community relationship is central throughout the research process, with community consultation being integral to data collection, data analysis, and the dissemination of results. For a book-length study of various projects in this vein, see Kimura and Kinchy (2019).

Defenders of the democratization view have mostly remained neutral about which of these methods are best for measuring public values. Evans, for example, prefers survey methods but writes that “deliberative events still fit with my overall point” (2023, 3). Schroeder suggests that deliberative minipublics offer “a promising starting point” (2021, 11), while also remaining open to citizen science, opinion surveys, and other methods (as do Alexandrova and Fabian 2022, see discussion on pp. 7–8). This methodological neutrality makes sense when the goal is to get the democratization view on the table. But ultimately, evaluating the view requires asking whether the data are adequate-for-purpose, and that, in turn, requires attention to the ways in which the data produced by these differing methods vary.

Some of the methods I have just outlined are good at producing data about what people value at a moment in time, without substantial opportunity for reflection, and without the provision of additional information that might lead them to revisit their initial opinions, values, and prioritizations. Preference elicitation and survey methods, and, to some extent, structured interviews, have this character.

Other methods are the opposite: They are good at producing data about what people value after they have learned new information, and after they have had discussions with others who have different viewpoints, as well as opportunities for iteration. Open-ended interviews, deliberative methods, and CBPR are like this.

So, the order in which I presented the different categories of methods roughly captures the way in which the data produced by these methods falls along a spectrum. At one end of the spectrum, the data capture people’s values that are less informed by any information beyond what subjects already have access to, while at the other end, the data reflect values that are more informed by external inputs.⁴ Of course, the data from within any methodological grouping may vary considerably along this axis as well. Surveys can provide more or less information about a topic, and deliberative methods can be more or less intentional about steering participants toward particular conclusions.

A natural thought at this point is that data from one end of the spectrum—the “informed by external information” end—are just better than data from the other end. Surely, we want data that reflect the public’s well-informed values rather than their snap judgments, for at least two reasons. First, our values are not bare normative preferences. Instead, we value what we value conditional on empirical beliefs (Hausman 2023). If there is reason, as there often is, to suspect that the public’s values depend on false beliefs, then it makes sense to use methods that may be able to correct these false beliefs (by providing information or facilitating discussion) *before* measuring values. Second, and echoing the discussion above, values data that are informed by external information can be contextually specific and therefore reduce the size of values gaps, that is, the gaps between values and value judgments that need to be bridged in order to say that science is being conducted “on the basis of” the public’s values. Indeed, for both of these reasons, there is a trend toward endorsing deliberative methods and CBPR across different literatures that deal with issues of public consultation and participation in science (Baker et al. 2021; DiMarco 2023; Fishkin 2009, 2020; Hausman 2023; Lusk 2021).

But a general preference for data generated via deliberative methods and CBPR is too quick. Yes, there is a lot to be gained by using methods that provide more information and more opportunities for interaction, deliberation, and iteration, but something is also lost. What is lost is *representativeness*. The more participants in a study go through an experience that changes the views they had at the outset, the less they can be viewed as a sample that represents the larger population from which they are drawn. As a result, the data about their values generated by the study are also less representative of the values held by the broader population.

In fact, it is central to the theory behind deliberative methods and CBPR that involvement does not just enable participants to articulate more clearly what they already think and value, but that it actually changes their minds. To the extent that participants' values are based on false empirical beliefs, the hope is that those values change when those beliefs are challenged. Considering the perspectives of other people (interaction) and returning to the same topic after time for reflection (iteration) are supposed to have the same effect. As Stanley Fishkin, a key proponent of deliberative methods in political science, writes, "...the thoughtful and informed views created in the experiment are not widely shared because the bulk of the public is still, in all likelihood, disengaged and inattentive" (2009, 28).

This loss of representativeness as a result of study methods that change their subjects' values in addition to measuring them has been termed a "transformative treatment" (Paul and Healy 2018). Such a study may provide counter-factual information about what the rest of the population from which the sample is drawn *would* value if they too had a similar experience, but it is not a good guide to what the population values absent such an experience. Cristina Lafont (2015) makes this point forcefully as well:

Deliberative Poll participants are supposed to be representative of the citizenry at large, in the *descriptive* sense of the term. Unfortunately, as we saw before, this is true if at all only regarding the "raw" opinions they have before the deliberative process, since the whole point of going through that process is to elicit a transformation of their initial judgments into qualitatively different judgments that are *for this very same reason* no longer representative of the actual views of the citizenry. (52)

What we have, then, is a tradeoff between different features of the data about public values. To the extent that research methods involve providing participants with more information, interaction, and iteration, their values become more contextually specific, more well-informed, and the data reflect this. At the same time, these data are less representative of the values of the broader population, because the research methods have caused the population and the sample to diverge. How to respond to this tradeoff? Which data should we prefer? There is no general answer to that question. Instead, the answer depends on the research purposes (Bokulich and Parker 2021). That means we need to get clearer on the purpose for which the democratization view means to use data about public values.

18.3 Understanding the research purpose

Earlier I said that different versions of the democratization view all share the idea that social scientific data are adequate for the purpose of legitimizing scientific conclusions. Now let's examine the idea of legitimacy more carefully. There are two elements in the democratization view's conception of legitimacy, and thus, two different conditions the data need to satisfy in order to be adequate for purpose. First, the values at play in value-laden science need to be the public's values. This is the *representativeness* condition. Second, the value judgments scientists make need to be appropriately based on the public's values. This is the *constraint* condition. So, in order for it to be the case that the social scientific data are adequate for the purpose of legitimizing scientific conclusions, they need to

be representative of the public's values while also providing sufficient constraints on scientists' value judgments. Ultimately, I will argue that the social scientific data can't really do this: The better they are at satisfying one of these conditions, the worse they are at satisfying the other, perhaps not in a strict, zero-sum sense, but there is a tradeoff, nonetheless. But before getting there, I will say more about each condition in turn.

The idea that value-laden science is legitimate when the values in question are the public's values comes from political philosophy. Drew [Schroeder \(2021\)](#) writes,

If democratic procedures or methods were carried out properly, then my values were an input into the process that yielded democratic values. My values are, in a sense, represented in the output of that process. This, at least according to many political theorists, means that those values should have a kind of legitimacy for me. (13)

Anna Alexandrova and Mark Fabian make a similar point when they say that “the process of specifying the context of a thick concept takes into account the relevant value judgments of those who lives stand to be affected by this research” (2022, 6). John Evans, too, endorses representativeness as an element of legitimacy: “if it is true that the principles in human research ethics actually do represent the primary values of the public...then this system also has a modicum of democratic legitimacy” (2023, 2).

In a democratic election, what makes the results legitimate, what gives them normative force, is that every citizen has the opportunity to vote. My preferred candidate may not win, but the election results still represent my values, and the values of every other voter, in the sense of being constituted by our votes. The democratization view claims that something similar can happen in science, even though the process by which our values are solicited and aggregated is different. Despite the methodological difference (social science, not voting), the legitimacy of the scientific results is secured by the fact that it is the public's values, and not the values of individual scientists, that are appealed to when it comes time to make unforced choices during scientific research. Legitimacy does not depend on me or any other individual person getting what we want, but it does require that *our* values be the starting point from which scientific value judgments are fashioned.

It also requires more than that. If the public's values are merely the starting point, but scientists are free to take them into account however they like, as anything from weak suggestions to strong prescriptions, then the democratization view collapses into views that allow scientists to pick and choose whatever social and political values seem best to them individually. That is why the second element of legitimacy involves the values of the public constraining the conclusions science can reach. Again, from Andrew Schroeder: “if scientific conclusions are based on objective facts plus democratic values, any two scientists investigating the same problem in the same social and political context should reach roughly the same conclusion” (2021, 11).⁵

One concern with this way of putting things is that it elides the distinction between values and value judgments, thus making it appear that scientific conclusions emerge easily and additively from facts plus values. In reality, it is more accurate to say that scientific conclusions are the products of empirical facts and value judgments. Terminology aside, however, the idea is that even allowing for some discretion in the way scientists bridge the gap between value and value judgment, once the public's values are known, scientists cannot easily reach divergent conclusions from the same set of facts and values.

There are two related and important motivations behind this element of legitimacy. First is the idea that the public's values cannot be made to support whatever judgment calls scientists are already disposed to make. In order for democratization to be meaningful, it has to make a real difference to the results that scientists report. Second is the idea that scientific objectivity requires something like

freedom from researcher idiosyncrasy (Boullicault and Schroeder 2021). We don't want the result of a study to depend on which scientist carried it out. But if the process of interpreting the public's values and converting them into value judgments is too flexible, then democratization does little to curb the problem of idiosyncrasy, and thus is insufficient to support objectivity.

18.4 Assessing the fit between data and research purpose

We now understand both the nature of the data available on public values and the purpose for which the democratization view would use these data. This puts us in a position to assess whether the data are adequate for purpose. More specifically, we are in a position to see the problem. In order for the representativeness condition to be met, the data on the public's values that scientists appeal to needs to be representative of the actual views of the public, not only of the sample that participated in studies of their values. But in order for the constraint condition to be met, we want data that minimizes values gaps, because it is in the process of bridging values gaps that scientists are likely to appeal to their own values, and to reach diverging conclusions despite having access to the same data about the public's values. As we saw in Section 18.2, there is a tradeoff between these two features of the data. The more representative it is, the less it carries with it additional information that would minimize values gaps, and vice versa.

This problem will not arise for all scientific value judgments, but my suspicion is that it will arise often enough to undermine the idea that social scientific data are adequate for the purpose of legitimizing scientific conclusions. It is a general and pervasive problem because of the ubiquity of values conflicts in science. By *values conflicts* I mean cases where there are multiple important values at play, and any actual value judgment will further one at the expense of others. Consider just the following two values, both of which we have good reason to think are important to the public: Preserving biodiversity and preventing species from moving outside of their "native" ranges. These two values come into conflict in a range of cases, though I will focus on just one here: Assisted migration research. Assisted migration is the term for management strategies in which humans move species or populations into new areas, sometimes within and sometimes beyond their native ranges, in order to help them respond to the impacts of climate change. Though it may be the only viable option for saving some species from extinction, assisted migration is controversial among ecologists because of the risk that it could disrupt the ecosystems to which threatened species are moved, potentially doing more harm than good.

Ecological research on assisted migration, then, involves making value judgments about how much evidence is enough evidence to conclude that a particular migration project will not have harmful unintended consequences. Depending on which values are considered more important in a particular case, a scientist may be more or less conservative in making this judgment. On one hand, the more important it is to try to save a given species from extinction, the more reason to set a lower evidential threshold for concluding that assisted migration is safe enough. On the other hand, the more important it is to avoid the risk of creating an invasive species, or otherwise changing the character of the transplant ecosystem, the more reason to set a higher evidential threshold.

The democratization view would have scientists look to the public's values in order to make such value judgments. So let's imagine how that would go. Merely knowing that both preserving biodiversity and preventing species from moving outside of their native ranges are important public values is insufficient to guide the value judgment. Scientists need information about the relative importance of these two values. And presumably, how important these two values are relative to one another will, for most people, vary depending on the context. For example, which species is at risk of extinction? What sort of ecosystem would it be transplanted to? In order to really get a sense of how the public thinks the value conflict should be resolved, then, an information-rich, and perhaps also deliberative

and iterative data collection process seems ideal. A deliberative event could provide education about assisted migration, the risks associated with it, the current state of the evidence about it, and a range of different applications. The resulting data would offer a nuanced picture of how participants navigate value conflict across a range of cases, which researchers could then apply to their own research contexts with minimal guesswork.

But of course, such a data collection process likely sacrifices representativeness, for the reasons I have already discussed. Learning tends to be transformative. The alternative, then, is to try to design a study that solicits general, and less contextual, information about the relative importance of the conflicting values. The resulting data would be more representative, but scientists would have to do a lot more work on the back end to make value judgments on the basis of this information. And the worry is that different scientists could reasonably come to opposite conclusions about the evidential threshold value judgment on the basis of the same social scientific data about the public's values.

Perhaps there is a way to avoid this tradeoff. Could a survey or interview simply ask participants how they think particular value judgments should be handled, providing barebones information about the decision scenario, thus avoiding the risk of a transformative treatment but also getting specific enough information to constrain scientific decisions? Maybe so. But recall that this strategy, as Kevin Elliott (2022) has pointed out, comes with strong feasibility concerns. The number of value judgments scientists make is just too large for it to be practical to survey the public about each one. At best we have added a third way in which scientific value judgments can fall short of legitimacy: Only a few will be able to live up to this standard.

18.5 Toward alignment

The question I have explored here is: Are social scientific data adequate for the purpose of making scientific conclusions legitimate in the way the democratization view claims? We have now arrived at an answer: No, because data that are adequate for constraining scientific conclusions will often not be representative of the broader public, and vice versa. Or, to the extent that it is possible for data to do both, the process of collecting this data will be very labor-intensive, making it impossible to base all or even many of science's unforced choices on public values.

I don't think this outcome means that the democratization view is wrong or that it is a bad way of thinking about what appropriate and responsible value judgments in science look like. The upshot, rather, is that the legitimizing purpose of the democratization view needs to be adjusted in light of what we know about the characteristics of the available data (Watkins 2024). There is more than one way of doing this, and investigating the different options is a task for future work in this area. One possible direction has already been partially explored by Greg Lusk (2021), who sides with an influential group of political theorists in preferring deliberative democracy to aggregative democracy as a model for the democratization view. In taking this position, Lusk seems to give up on the representativeness condition in favor of constraint (though for different reasons than I have given here), and in doing so may avoid the adequacy-for-purpose issues that I have raised.

Regardless of how the proponents of the democratization view decide to respond to the adequacy-for-purpose challenge, the future of the view requires a closer engagement with social science in order to make sure that its normative and empirical dimensions are aligned with one another.

Notes

- 1 This is by now a consensus view in philosophy of science; see, e.g., Rudner (1953), Douglas (2009), and Elliott (2017). I adopt the terminology of unforced choices from Matthew Brown (2020).
- 2 A version of this view has been defended by, among others, de Melo-Martín and Intemann (2018); Schroeder (2021); Lusk (2021); Alexandrova and Fabian (2022); DiMarco (2023); Evans (2023).

- 3 Throughout the chapter, I will rely on the adequacy-for-purpose view of data developed by Bokulich and Parker (2021), and a recent refinement of the view proposed by Watkins (2024).
- 4 There are also well-known criticisms of each of these families of methods, so to some extent it is contentious to say that any of these methods is “good at” producing certain kinds of results. Here, I set such controversies aside in order to focus on a different set of issues.
- 5 See also Evans (2023) and Alexandrova and Fabian (2022) for similar discussions of the importance of constraint.

References

- Alexandrova, Anna, and Mark Fabian. 2022. “Democratising Measurement: Or Why Thick Concepts Call for Coproduction.” *European Journal for Philosophy of Science* 12 (7): 1–23. <https://doi.org/10.1007/s13194-021-00437-7>.
- Baker, Rachel, Helen Mason, Neil McHugh, and Cam Donaldson. 2021. “Public Values and Plurality in Health Priority Setting: What to Do When People Disagree and Why We Should Care about Reasons as Well as Choices.” *Social Science & Medicine* 277: 113892. <https://doi.org/10.1016/j.socscimed.2021.113892>.
- Bokulich, Alisa, and Wendy Parker. 2021. “Data Models, Representation and Adequacy-For-Purpose.” *European Journal for Philosophy of Science* 11: 1–26. <https://doi.org/10.1007/s13194-020-00345-2>.
- Boulcault, Marion, and S. Andrew Schroeder. 2021. “Public Trust in Science: Exploring the Idiosyncrasy-Free Ideal.” In *Social Trust*, edited by Kevin Vallier and Michael Weber. Routledge.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- de Melo-Martin, Immaculada, and Kristen Intemann. 2018. *The Fight against Doubt: How to Bridge the Gap between Scientists and the Public*. Oxford University Press.
- DiMarco, Marina. 2023. “Cooperative Epistemic Trustworthiness.” *Philosophy of Science* 90 (5): 1100–09. <https://doi.org/10.1017/psa.2023.61>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin C., and Rebecca Korf. 2024. “Values in Science: What Are Values, Anyway?” *European Journal for Philosophy of Science* 14: 1–24. <https://doi.org/10.1007/s13194-024-00615-3>.
- Evans, John H. 2023. “Translational Bioethics and Public Input.” *Ethics & Human Research* 45 (4): 35–39. <https://doi.org/10.1002/eahr.500175>.
- Evans, John H., and Cynthia E. Schairer. 2024. “Scientism, Trust, Value Alignment, Views of Nature, and US Public Opinion about Gene Drive Mosquitos.” *Public Understanding of Science* 33 (7): 884–901. <https://doi.org/10.1177/09636625241229196>.
- Fishkin, James S. 2009. *When The People Speak: Deliberative Democracy and Public Consultation*. Oxford University Press.
- Fishkin, James S. 2020. “Cristina Lafont’s Challenge to Deliberative Minipublics.” *Journal of Deliberative Democracy* 16 (2): 56–62. <https://doi.org/10.16997/jdd.394>.
- Hausman, Daniel M. 2023. “Eliciting Preferences and Respecting Values: Why Ask?” *Social Science & Medicine* 320: 115711. <https://doi.org/10.1016/j.socscimed.2023.115711>.
- Holman, Bennett, and Torsten Wilholt. 2022. “The New Demarcation Problem.” *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Kimura, Aya H., and Abby Kinchy. 2019. *Science by the People: Participation, Power, and the Politics of Environmental Knowledge*. Rutgers University Press.
- Lafont, Cristina. 2015. “Deliberation, Participation, and Democratic Legitimacy: Should Deliberative Mini-Publics Shape Public Policy?” *Journal of Political Philosophy* 23 (1): 40–63. <https://doi.org/10.1111/jopp.12031>.
- Lusk, Greg. 2021. “Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere.” *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Murray, Christopher J., and A. Lopez. 1996. *The Global Burden of Disease*. Harvard University Press.
- Paul, Laurie A., and Kieran Healy. 2018. “Transformative Treatments.” *Nous* 52 (2): 320–35. <https://doi.org/10.1111/nous.12180>.

- Resnik, David B., and Kevin C. Elliott. 2023. "Science, Values, and the New Demarcation Problem." *Journal for General Philosophy of Science* 54 (2): 259–86. <https://doi.org/10.1007/s10838-022-09633-2>.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2). <https://doi.org/10.1093/bjps/axz023>.
- Stange, Erik, Dagmar Hagen, Berit Junker-Köhler, and Bjørn P. Kaltenborn. 2022. "Public Perceptions of Ecological Restoration within the Context of Norwegian Landscape Management." *Restoration Ecology* 30 (7): e13612. <https://doi.org/10.1111/rec.13612>.
- Ward, Zina B. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science Part A* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Watkins, Aja. 2024. "The Adequacy of Purposes for Data: A Paleocological Case Study." *Synthese* 203: 159. <https://doi.org/10.1007/s11229-024-04588-x>.

FDA EVIDENTIARY STANDARDS AND THE NEED TO ATTEND TO STAKEHOLDERS' VALUES

Inmaculada de Melo-Martín

19.1 Introduction

Contextual value judgments play important and beneficial roles in scientific reasoning (Douglas 2009; Elliott 2017; Longino 1990). Although it is neither possible nor desirable to prevent non-epistemic values from operating in science, legitimate concerns exist about whose values should be guiding research (de Melo-Martín 2024). Arguably, judgments about which values to endorse or incorporate in research should not be left to scientists alone as this would give them disproportionate power in shaping science that informs policy decisions (Betz 2013; Pielke 2007). Since scientists have no special expertise or authority in making these judgments and are unrepresentative of the diverse stakeholders affected by science, it is questionable that they alone should decide what values to use when conducting research. Because there can be some reasonable disagreements about social, political, and ethical values, in pluralistic societies all stakeholders should have some say in determining which values to endorse or prioritize in cases of conflict.

To address this value-imposition concern, many philosophers of science have called for identifying democratic and deliberative processes that facilitate that relevant parties, rather than just scientists, can determine what values should guide scientific research (Elliott 2017; Intemann 2015; Lusk 2021; Schroeder 2021). Often these proposals advocate engaging relevant stakeholders in various ways, including community-based advisory boards, citizen panels, deliberative polling from relevant communities, or seeking consensus (Douglas 2009; Elliott 2017; Intemann 2015; Schroeder 2021). Such engagement is especially relevant in the context of the biomedical sciences, where failing to attend to relevant stakeholders can have particularly negative implications (Porter 2022). Patients and their caregivers have direct experiences with illness and medical care: The specific challenges they face in their daily lives, the most concerning aspects of the disease, the interventions that would be more helpful, and the concessions they are willing to make in managing their disease. It therefore makes sense to involve them in various phases of clinical research (Domecq et al. 2014). For instance, clinical trial success critically depends on ensuring the participation of an adequately informed and appropriately diverse group of patients. Insufficient or skewed participation is a common problem in clinical research (Desai 2020). Trials might impose too high a burden for patients: include too many visits or interventions, have unnecessarily restrictive inclusion and exclusion criteria, be excessively lengthy, or involve unreasonable benefits-risks profiles from the patients' perspectives (Poongothai et al. 2023). These factors hinder participation and result in reduced statistical power,

inconclusive results, and misalignment between the internal and external validity of clinical trials (Poongothai et al. 2023). Collaborations with relevant patient groups can minimize these problems (Crocker et al. 2018). Importantly, these collaborations can challenge not only research priorities or product development but also scientific methodology (Epstein 1996).

Although it seems both ethically and epistemically appropriate to identify democratic procedures that would ensure that the values shaping research represent the interests of relevant stakeholders, doing so successfully faces serious practical and theoretical challenges (Le Bihan 2024). Here, I use the US Food and Drug Administration's (FDA's) attempts to incorporate stakeholders' values to show some of these challenges. Values play central roles in drug and medical device research. For instance, inductive risk judgments are central to drug development and approval processes, where decisions about how much evidence or what type of evidence is needed to determine safety or efficacy affect how quickly such evidence can be gathered—and thus, how fast drugs can arrive to the market. I point out challenges regarding the incorporation of what are arguably legitimate but conflicting values from relevant stakeholders. Certainly, many of those calling for public engagement as a way to address the value imposition concern have recognized challenges when stakeholders have a plurality of legitimate values (Elliott 2017; Intemann 2015; Kitcher 2011; Lusk 2021; Schroeder 2021). However, I believe that the implications of those challenges for such engagement proposals have not been sufficiently appreciated. I first offer a brief overview of the FDA's role in drug development and approval processes and its attempts to incorporate the legitimate values of relevant stakeholders. I then discuss several of the challenges such attempts face and conclude with some suggestions about how to move forward.

19.2 The FDA: Attending to stakeholders' values

The FDA is one of the most respected and powerful public agencies in the world (Carpenter 2010). As of October 2024, the FDA was responsible for the oversight of more than \$3.9 trillion worth of food, tobacco, and medical products produced in the US and abroad (Food and Drug Administration (FDA) 2024). Only drugs explicitly declared safe and effective for their intended uses by the FDA can be legally marketed in the US. Hence, although the FDA's formal authority applies only to the US and its territories, its decisions have global reach, as any international drug company wishing to introduce products to the US market is required to obtain FDA approval. Moreover, many other regulatory agencies all over the world have used the FDA as a model for their approval processes (Gaudillière and Hess 2013).

After its foundation in 1906, the FDA gradually increased its gatekeeping powers over pharmaceutical markets (Carpenter 2010). It originally verified drug composition through laboratory tests to prevent adulteration and fraud. In 1938, after the sulfanilamide elixir tragedy, Congress passed the Federal Food, Drug, and Cosmetic Act, which gave the FDA authority to require drug manufacturers to prove the safety of their products (Carpenter 2010). A new pharmaceutical catastrophe—thalidomide—led Congress in 1962 to pass the Kefauver-Harris Drug Amendments. For the first time, drug manufacturers were required to show not only the safety of their products but also their efficacy. In particular, sponsors were required to provide “substantial evidence” of effectiveness based on well-controlled clinical studies. The Kefauver-Harris amendments also required FDA approval before a drug could be marketed in the US.

Until recently, the FDA's evidentiary standard for allowing a new drug onto the market was, with some exceptions, two double-blind, randomized-controlled trials (RCT). However, although the RCT is usually considered the gold standard of evidence, it has important drawbacks from the perspective of at least some patients (Epstein 1996; Will and Moreira 2010). RCTs incorporate debiasing methods to secure a like-to-like comparison: Allocation of the intervention is randomized, and the

interventions are masked (Hackshaw 2009). These debiasing mechanisms, however, impose various costs on patients (Epstein 1996; Will and Moreira 2010). First, patients who participate in trials perceive randomization, masking, and the use of placebos as contrary to their interests. This is particularly the case when studying interventions for life-threatening illnesses. Second, trial protocols often have restrictive inclusion and exclusion criteria, limiting the type of participants who can have access to them. Third, protocols are also quite strict, and many participants find it difficult to adhere to them and drop out before completing the study. Furthermore, due to complexities of design and conduct, RCTs can take a long time to complete, making it difficult for patients in need to access new treatments.

Patient groups have argued for the need to take into account their interests in two important ways (Epstein 2007; Flanigan 2017; Wehling, Viehöver, and Koenen 2015). First, they have insisted on a change of evidentiary standards for drug development and approval in ways that make trials less burdensome and bring innovations to the market faster. Second, they have demanded that their preferences regarding what might be onerous about a disease, how to treat it, or what factors to address should be considered when making decisions about what interventions to develop and approve. Attempting to incorporate stakeholders' values regarding these two concerns—to accomplish faster access to drugs and to ensure attention to patients' preferences regarding particular interventions—present challenges. Because of space limitations, I will limit my discussion to the issue of faster access to medications. Of course, some of the challenges of attending to these two concerns will be common, but some are specific to each of them. In particular, incorporating some stakeholders' values in the context of speedier access to drugs has direct effects on standards of evidence for safety and effectiveness.

Since the 1980s, the FDA, responding to congressional demands, has attempted to address patients' concerns about faster and easier access to new treatments. It has implemented several expedited programs that allow drugs to be approved provisionally with reduced evidentiary requirements (Darrow, Avorn, and Kesselheim 2020). For instance, the Fast-Track program permits approval of new interventions for serious and life-threatening conditions without conducting phase 3 trials, and the Accelerated Approval program allows the use of surrogate measures only reasonably likely to predict clinical benefit as endpoints for the pivotal clinical trials that form the basis for approving a new drug.

These efforts have been further supported by several provisions in the *21st Century Cures Act* (2016), which aim to expedite clinical testing and approval of new drugs and indications. Importantly, these provisions alter the types of data that pharmaceutical and device companies can use to obtain FDA approval or to meet post-approval requirements. In particular, one of the sections (Sec. 3021) directs the FDA to consider the use of complex innovative trial designs (CID) in the development and regulatory decision-making of drugs and biological products (Food and Drug Administration (FDA) 2020). Although the FDA does not offer a definition of CID, it takes them to include trial designs that have rarely or never been used to date to provide substantial evidence of effectiveness in new drug or biologics applications. Some examples are trial designs that formally borrow external or historical information or control arm data from previous studies to expand upon concurrent controls.

Another section of the Cures Act (Sec. 3022) supports the use of “real world evidence” in the approval of a new indication of an already approved drug or to satisfy post-approval requirements. The Act defines real-world evidence (RWE) as data regarding the usage, or the potential benefits or risks of a drug derived from sources other than randomized clinical trials. Such sources include healthcare information from electronic health records, billing databases, product and disease registries, patient input, and observational studies. Collecting these types of data is frequently less onerous than conducting an RCT and they can be collected in situations where RCTs are unfeasible. Moreover, RWE draws upon various data sources that incorporate the experiences of broader patient cohorts. Patient

groups, and other stakeholders, lobbied Congress regarding these provisions and public comment documents on the draft of the Bill show that patient groups expressed support for these changes to evidentiary standards (Hwang, Sachs, and Kesselheim 2017).

19.3 Incorporating stakeholders' values: Challenges

As mentioned earlier, the recognition that contextual values play relevant roles in scientific research has led to concerns about the disproportionate power that scientists could have in shaping research. Strategies to ensure incorporation of relevant stakeholders' values offer a way to address these concerns. But such incorporation presents at least two interrelated challenges: What to do when a relevant stakeholder has conflicting values and how to address conflicts of values among pertinent stakeholders.

Current patients are clearly relevant stakeholders in decisions about drug development and approval. They have interests in ensuring fast access to medications, particularly when drugs can be lifesaving or improve their quality of life in significant ways. When they participate in clinical trials, they also have an interest in the use of evidence that is less onerous. RWE and surrogate endpoints, for instance, often require less time to collect than traditional clinical outcome data. These alternative measures are also helpful when dealing with conditions that have small patient populations and where enrolling enough participants in traditional phase 3 trials might be difficult (Wedam et al. 2020). Likewise, existing patients with few other options to treat their diseases might have strong preferences that are inconsistent with masking, randomization, or the use of placebos (Ali et al. 2021). The FDA's regulations allowing sources of evidence other than those produced by RCTs are a way to attend to these legitimate interests.

However, although patients want faster access to needed drugs, they also want them to be safe and effective.¹ Some evidence indicates that these different legitimate interests can conflict. For example, accelerated approval is now common in oncology, with about one-third of all oncology drug approvals using the pathway (Scott et al. 2023). The surrogate measures that are usually used for accelerated approvals of cancer treatments include tumor response rate and progression-free survival. But these surrogate endpoints correlate poorly with overall survival (Gyawali, Phillips Hey, and Kesselheim 2020). A significant amount of evidence shows that most of the drugs approved by this pathway ultimately fail to demonstrate benefit in overall survival or quality of life in confirmatory trials (Gyawali, Phillips Hey, and Kesselheim 2019; Liu, Kesselheim, and Scheffer Cliff 2024). For many of these drugs, the FDA has requested withdrawals because of the lack of benefit (Preziosi and Priefer 2024). These problems also occur with non-oncology drugs (Omae et al. 2022).

It is true that the FDA usually requires post-approval confirmatory studies to validate true clinical benefit to either convert the drug to standard approval or remove it from the market. Such trials should ideally be conducted in a reasonable timeframe to limit harms to patients and should use meaningful clinical endpoints rather than unvalidated surrogate ones. However, evidence shows that post-approval trials are often delayed, in some cases for years after approval (Mao, Alexander, and Li 2024). Furthermore, when conducted, many of the confirmatory trials continue using surrogate measures, sometimes the same surrogate points used for accelerated approval. Additionally, medications approved by special pathways are often integrated into clinical practice before confirmation of benefit for original indications (Naci et al. 2017). This means that, for years, patients are exposed to drugs that might have no clinical benefit and could have significant adverse effects, waste time and money, and are prevented from pursuit of other alternatives, potentially including palliation (Lynch and Bateman-House 2020; Mostaghim, Gagne, and Kesselheim 2017).

This conflict is highlighted when one considers another stakeholder in the drug development and approval processes: Drug manufacturers. Although drug manufacturers arguably have an interest in

bringing safe and effective products to the market, as publicly traded corporations, they are structured to value shareholder interests over those of patients when those interests clash. Their main mission is to sell profitable drugs, whether or not those drugs are optimally useful in treating diseases, and a variety of market and regulatory mechanisms incentivize pharmaceutical companies' attention to profits over the health of patients (Madl 2019; Torreele 2024). Importantly, the interests of patients in obtaining faster access to effective drugs serve drug manufacturers' interests well. Pharmaceutical companies also have an interest in getting innovative products to the market as fast as possible, as this increases profits. For the same reason, they have an interest in cutting development costs. Accelerated approvals are consistent with these interests. Given that companies benefit financially from accelerated approvals, this creates an incentive for them to engage with patient groups that also seek faster drug approval. Indeed, some evidence suggests that the interests of patients can be co-opted in ways that have been underappreciated (Holman and Geislar 2018; Tempini and Teira 2019). After all, drug manufacturers face few negative consequences when obtaining accelerated approval for their drugs.

This conflict in values—faster access to possibly beneficial treatments vs more robust evidence of safety and efficacy—is at the heart of the FDA Accelerated Approval program as well as other special pathways that aim to bring needed therapies to patients faster. Although these programs can be modified to help better balance these divergent patient's interest (Andreoletti and Blasimme 2023; Fashoyin-Aje et al. 2022), simply calling for inclusion of patients' values when making these decisions will not solve the problem, as patients' values in this context are actually in conflict. Hence, to the extent that patients have a legitimate interest in maintaining standards of evidence more likely to produce safe and effective treatments, this would impose limits on how fast treatments can be brought to the market. The difficulty in managing this conflict can even be observed in the Cures Act mandates. Although Congress asks the FDA to use RWE, novel clinical trials, and patient experience data in the approval process to get drugs to patients faster, it also stipulates that the various amendments the Cures Act makes to the Federal Food, Drug, and Cosmetic Act are not intended to alter the "substantial evidence standard."

Current patients, however, are not the only relevant stakeholders when it comes to the incorporation of values in drug research and approval. Arguably, citizens and healthcare providers are also appropriate stakeholders whose interests should be considered when making decisions about drug development and approval.

And although the interests of these various stakeholders can converge, conflicts are also likely. For instance, in a controversial decision in 2021, the FDA granted accelerated approval to aducanumab, a drug indicated for the treatment of Alzheimer's disease (AD) (Alexander et al. 2021). The approval was based on an unvalidated surrogate outcome: Reduction in beta-amyloid plaques. Many considered this evidence insufficient as proof of clinical benefit and aducanumab was also associated with significant risks, including brain swelling and hemorrhage in more than one-third of patients who received the dose approved by the FDA (Rabinovici 2021). According to some evidence, Alzheimer's patient groups pushed for the approval of the drug (Hu 2023). This seems understandable given that AD is a devastating disease that leads to irreversible cognitive and functional decline and one for which no drug had ever been approved for modification of its course. Indeed, the FDA emphasized that patients and their families were willing to accept the uncertainty and risks associated with aducanumab treatment in exchange for earlier access to a potentially effective drug (Dunn, Stein, and Cavazzoni 2021).

But, although the accelerated approval of aducanumab serves the interests of patients suffering from AD in accessing a drug that might slow the disease, the interests of other stakeholders are arguably also relevant. Citizens who are not yet but might become future patients also have an interest in ensuring that drugs are brought to the market only after robust evidence of safety and efficacy. This makes the special approval pathways concerning because, as mentioned earlier, evidence shows that

confirmatory trials fail to demonstrate benefit for many of the drugs approved by these pathways (Gyawali, Phillips Hey, and Kesselheim 2019; Liu, Kesselheim, and Scheffer Cliff 2024), with many of them being withdrawn because of the lack of benefit (Omae et al. 2022; Preziosi and Priefer 2024). In some cases, withdrawal from the market can happen relatively quickly, minimizing harms to patients. This was the case with aducanumab. In January 2024, Biogen, the maker of aducanumab announced that it would stop developing and selling the drug. After its accelerated approval in 2021, Biogen launched an international post-marketing study designed to evaluate its real-world safety and effectiveness in US clinical practice. However, the trial was terminated in May 2022 because it failed to meet clinical projections (Heidebrink and Paulson 2024). Not uncommonly, however, confirmatory trials are delayed, and when conducted, they may suffer from limitations similar to those of preapproval studies (Woloshin et al. 2017). Hence, drugs can be in the market for years despite uncertain clinical benefit and lack of information about adverse effects, putting patients at risk of harm. Furthermore, once the drugs are in the market, they are integrated into clinical practice and used for off-label indications, which expands their use even when evidence of efficacy for original indication has not been confirmed (Naci et al. 2017).

Citizens also have an interest in ensuring affordability and appropriate use of their tax money. Accelerated approved drugs often have extremely high costs. Biogen, for instance, initially charged \$56,000 yearly (Chiong et al. 2022). This price did not include costs related to infusion services, the necessary fees for increased physician follow-up, or the additional studies needed to monitor for risks associated with use of the drug. The price of cancer drugs receiving accelerated approval can be even higher, with many of them priced at more than \$100,000 per year before confirmatory trials have been completed (Frank, Shahzad, and Emanuel 2022).

This, of course, presents problems for individuals who need to use the drugs in question. But it constitutes a problem also for citizens in general. For example, Medicare and Medicaid spend millions of dollars covering the costs of accelerated approval drugs and beneficiaries of these programs spend millions of dollars in out-of-pocket spending on these drugs (Shahzad, Naci, and Wagner 2021; Skydel et al. 2022). A recent study of 38 drugs that received FDA accelerated approval from 2012 to 2017 found that the Centers for Medicare & Medicaid spent nearly \$68 billion through 2020 (Skydel et al. 2022). Importantly, 59% of the spending was for drugs with confirmatory trials that only evaluated surrogate endpoints. Likewise, Medicare has paid millions of dollars for drugs that were either voluntarily withdrawn by the manufacturers or recommended by the FDA for withdrawal (Shahzad, Naci, and Wagner 2021). Furthermore, when drugs granted accelerated approval are placed on treatment guidelines, they are then covered by private insurance providers (Rodriguez et al. 2021). This clearly benefits patients who need the drugs but also exposes them to high out-of-pocket costs and importantly also exposes others to higher insurance premiums. Increased out-of-pocket expenses are not only a financial burden for patients; they also often reduce levels of patient compliance and lead to unfavorable outcomes (Gellad and Kesselheim 2017). Ultimately, the high costs of drugs for which clinical benefit is uncertain and sometimes nonexistent put burdens on public services that are unsustainable.

Healthcare providers are also relevant stakeholders whose values should be considered in decisions regarding drug development and approval processes. They have a fiduciary duty to prioritize patients' health and wellbeing and also a duty to nurture the trust of their patients. Insofar as drugs approved by accelerated pathways are incorporated into clinical practice before sufficient evidence of safety and efficacy, clinicians run the risk of prescribing medications that are harmful to their patients. Harms can be the result of unknown side effects, the use of medications that are ineffective and whose use prevents patients from taking other measures, or because of wasted financial resources. But prescription of medications with uncertain safety and benefit profile risks undermining patients' warranted trust in clinicians.

Ensuring that patients have appropriate information regarding uncertainties in this context can minimize some of the harms mentioned. Providing patients with appropriate information is, however, challenging for clinicians. First, evidence suggests that clinicians often lack familiarity with drug and medical device regulatory practices, believe that the evidence supporting FDA approvals is more rigorous than it often is, and are unaware when medical products have expedited approvals (Dhruva, Kesselheim, and Woloshin 2024). In many cases, labels for medications that have received accelerated approval fail to include this information or to describe the surrogate marker(s) that supported the accelerated approval (Ballreich et al. 2023). Second, accelerated approvals increase decisional burdens on clinicians who need to ensure robust discussions regarding the ambiguity of clinical benefits for drugs approved by special pathways and must help patients make risk-benefit decisions in situations of greater uncertainty. Such decisions are particularly difficult in cases like that of aducanumab because clinical trials were very selective, excluding participants with comorbidities that are common in the AD population (Anderson et al. 2021). These exclusions make it harder for clinicians to determine when the risk-benefit balance is reasonable when advising patients suffering from conditions that were excluded because of the possibility of increased risks. Indeed, some studies have shown that many clinicians were unlikely to prescribe the drug (Dhruva et al. 2023).

Considerations about familiarity with FDA regulatory practices and the additional time spent with patients trying to determine whether particular drugs approved by special pathways are appropriate are arguably important when determining accelerated approvals. Likewise, questions about how clinicians should deal with the burdens on patients associated with prescribing drugs of unproven benefit but with high costs that are either not covered by insurance or that involve elevated out-of-pocket expenses, and concerns about undermining trust in the doctor-patient relationship, are also relevant when making decisions about drug development and approval processes. The interests of clinicians can conflict with those of patients who seek faster access to needed medications. More challenging still is that clinicians are confronted with the conflicting interests of current patients who want access to novel therapeutics as fast as possible and those of future patients who would benefit from drug approvals with less uncertainty about their risks and potential clinical benefits.

19.4 Concluding remarks

Incorporating the interests of various stakeholders into scientific research seems an appropriate goal both on epistemic and ethical grounds. Nonetheless, the FDA's attempts to do so illustrate the challenges involved. Not uncommonly, debates about engaging publics to incorporate values in science presuppose that some values are clearly legitimate and should be included, while others are illegitimate and should be excluded. Those cases might still be challenging as disagreements might actually exist about what the legitimate and illegitimate values in fact are (de Melo-Martín 2024).

Here, however, I have called attention to what I believe is a harder challenge to address. In the context of drug development and approval processes, the FDA is charged with attending to various values of relevant stakeholders that arguably are all legitimate but that conflict. While current patients suffering from diseases with few options have an interest in faster access to drugs and in less burdensome means of gathering evidence, they ultimately also want drugs that will help them with their diseases, and this might require more robust evidentiary standards. Future patients have interest in ensuring that drugs are marketed only after sufficient evidence of safety and efficacy. Clinicians want to ensure that they prescribe drugs that help their patients. Citizens have an interest in an appropriate use of their tax dollars. Attending to some of these values necessarily calls for putting other legitimate stakeholders' values aside.

Of course, decisions about managing these conflicts are possible. After all, we manage conflicts among legitimate values in many other contexts such as funding priorities. But calling for the incorporation

of stakeholders' values in science does not, by itself, constitute a solution. Certainly, bringing relevant stakeholders together might provide relevant information about the values at stake and can allow them to deliberate (Intemann 2015; Schroeder 2021), even if it would still be the case that in some instances, it will simply not be possible to incorporate all legitimate values of all relevant stakeholders. Proposals for incorporating legitimate stakeholder values have focused primarily in proposing mechanisms to determine which or whose values ought to be incorporated into scientific research (Intemann 2015; Schroeder 2021), under the assumption that not all values are legitimate. When all the values are legitimate, these proposals are insufficient. More attention needs to be given to proposing strategies to manage the legitimate but conflicting values of relevant stakeholders (Laursen et al. 2021). Likewise, the existence of a plurality of legitimate values might also be a reason for researchers and funders to be more open to conducting and supporting research shaped by alternative values (Hillgardt 2023). Although this option will require difficult decisions about resource allocation, at least in some cases it might be the only way to ensure that important interests are not neglected when conducting research.

Acknowledgments

This work was supported by the National Center for Advancing Translational Science of the National Institute of Health under Grant UL1TR002384.

Note

- 1 Depending on their circumstances, individuals might give more weight to some of these interests than others, but they clearly have these various, conflicting, interests.

References

- 21st Century Cures Act. H.R. 34, 114th Congress. 2016. <https://www.gpo.gov/fdsys/pkg/BILLS-114hr34enr/pdf/BILLS-114hr34enr.pdf>.
- Alexander, G. Caleb, David S. Knopman, Scott S. Emerson, et al. 2021. "Revisiting FDA Approval of Aducanumab." *New England Journal of Medicine* 385: 769–71. <https://doi.org/10.1056/NEJMp2110468>.
- Ali, Samina, Gareth Hopkin, Naveen Poonai, et al. 2021. "A Novel Preference-Informed Complementary Trial (PICT) Design for Clinical Trial Research Influenced by Strong Patient Preferences." *TRIALS* 22 (1). <https://doi.org/10.1186/s13063-021-05164-1>.
- Anderson, Timothy S., John Z. Ayanian, Jeffrey Souza, and Bruce E. Landon. 2021. "Representativeness of Participants Eligible to Be Enrolled in Clinical Trials of Aducanumab for Alzheimer Disease Compared with Medicare Beneficiaries with Alzheimer Disease and Mild Cognitive Impairment." *JAMA* 326: 1627–29. <https://doi.org/10.1001/jama.2021.15286>.
- Andreoletti, Mattia, and Alessandro Blasimme. 2023. "Accelerated Drug Approval: Meeting the Ethical Yardstick." *Bioethics* 37: 647–55. <https://doi.org/10.1111/bioe.13191>.
- Ballreich, Jeromie, Mariana Socal, Charles L. Bennett, Andrew Xuan, Antonio Trujillo, and Gerard Anderson. 2023. "Accelerated Approval Drug Labels Often Lack Information for Clinical Decision-Making." *Pharmacotherapy* 43 (4): 300–4. <https://doi.org/10.1002/phar.2789>.
- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Carpenter, Daniel P. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press.
- Chiong, Winston, Benjamin David Tolchin, Richard J. Bonnie, et al. 2022. "Decisions with Patients and Families Regarding Aducanumab in Alzheimer Disease, with Recommendations for Consent AAN Position Statement." *Neurology* 98: 154–59. <https://doi.org/10.1212/WNL.0000000000013053>.
- Crocker, Joanna C., Ignacio Ricci-Cabello, Adwoa Parker, et al. 2018. "Impact of Patient and Public Involvement on Enrolment and Retention in Clinical Trials: Systematic Review and Meta-analysis." *BMJ* 363: k4738. <https://doi.org/10.1136/bmj.k4738>.

- Darrow, Jonathan J., Jerry Avorn, and Aaron S. Kesselheim. 2020. "FDA Approval and Regulation of Pharmaceuticals, 1983-2018." *JAMA* 323 (2): 164–76. <https://doi.org/10.1001/jama.2019.20288>.
- de Melo-Martin, Inmaculada. 2024. "Concerns about Contextual Values in Science and the Legitimate/Illegitimate Distinction." *Philosophy of Science* 91: 851–68. <https://doi.org/10.1017/psa.2024.20>.
- Desai, Mira. 2020. "Recruitment and Retention of Participants in Clinical Studies: Critical Issues and Challenges." *Perspectives in Clinical Research* 11: 51–53. https://doi.org/10.4103/picr.PICR_6_20.
- Dhruva, Sanket S., Aaron S. Kesselheim, Steven Woloshin, et al. 2023. "Physician Perspectives on the Food and Drug Administration's Decision to Grant Accelerated Approval to Aducanumab for Alzheimer's Disease." *Clinical Pharmacology & Therapeutics* 114: 614–17. <https://doi.org/10.1002/cpt.2954>.
- Dhruva, Sanket S., Aaron S. Kesselheim, Steven Woloshin, et al. 2024. "Physicians' Perspectives on FDA Regulation of Drugs and Medical Devices: A National Survey." *Health Affairs* 43 (1): 27–35. <https://doi.org/10.1377/hlthaff.2023.00466>.
- Domecq, Juan Pablo, Gabriela Prutsky, Tarig Elraiyah, et al. 2014. "Patient Engagement in Research: A Systematic Review." *BMC Health Services Research* 14. <https://doi.org/10.1186/1472-6963-14-89>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Dunn, Billy, Peter Stein, and Patrizia Cavazzoni. 2021. "Approval of Aducanumab for Alzheimer Disease – The FDA's Perspective." *JAMA Internal Medicine* 181: 1276–78. <https://doi.org/10.1001/jamainternmed.2021.4607>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. University of California Press.
- Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. University of Chicago Press.
- Fashoyin-Aje, Lola A., Gautam U. Mehta, Julia A. Beaver, and Richard Pazdur. 2022. "The On- and Off-Ramps of Oncology Accelerated Approval." *New England Journal of Medicine* 387: 1439–42. <https://doi.org/10.1056/NEJMp2208954>.
- Flanigan, Jessica. 2017. *Pharmaceutical Freedom: Why Patients Have a Right to Self-Medicate*. Oxford University Press.
- Food and Drug Administration (FDA). 2020. Interacting with the FDA on Complex Innovative Trial Designs for Drugs and Biological Products. Guidance for Industry. Available at <https://www.fda.gov/media/130897/download>.
- Food and Drug Administration (FDA). 2024. *FDA at a Glance*. Available at <https://www.fda.gov/media/182749/download>.
- Frank, Richard G., Mahnum Shahzad, and Ezekiel J. Emanuel. 2022. "Accelerated Approval of Cancer Drugs: No Economic Reward for Drug Makers That Conduct Confirmatory Trials." *Health Affairs* 41: 1273–80. <https://doi.org/10.1377/hlthaff.2022.00119>.
- Gaudillière, Jean-Paul, and Volker Hess. 2013. *Ways of Regulating Drugs in the 19th and 20th Centuries*. Palgrave MacMillan.
- Gellad, Walid F., and Aaron S. Kesselheim. 2017. "Accelerated Approval and Expensive Drugs—A Challenging Combination." *New England Journal of Medicine* 376: 2001–4. <https://doi.org/10.1056/NEJMp1700446>.
- Gyawali, Bishal, Spencer Phillips Hey, and Aaron S. Kesselheim. 2019. "Assessment of the Clinical Benefit of Cancer Drugs Receiving Accelerated Approval." *JAMA Internal Medicine* 179: 906–13. <https://doi.org/10.1001/jamainternmed.2019.0462>.
- Gyawali, Bishal, Spencer Phillips Hey, and Aaron S. Kesselheim. 2020. "Evaluating the Evidence behind the Surrogate Measures Included in the FDA's Table of Surrogate Endpoints as Supporting Approval of Cancer Drugs." *EClinicalMedicine* 21: 100332. <https://doi.org/10.1016/j.eclinm.2020.100332>.
- Hackshaw, Allan K. 2009. *A Concise Guide to Clinical Trials*. Wiley-Blackwell.
- Heidebrink, Judith L., and Henry L. Paulson. 2024. "Lessons Learned from Approval of Aducanumab for Alzheimer's Disease." *Annual Review of Medicine* 75: 99–111. <https://doi.org/10.1146/annurev-med-051022-043645>.
- Hillgardt, Hannah. 2023. "Partisan Science and the Democratic Legitimacy Ideal." *Synthese* 202: 135. <https://doi.org/10.1007/s11229-023-04370-5>.
- Holman, Bennett, and Sally Geislar. 2018. "Sex Drugs and Corporate Ventriloquism: How to Evaluate Science Policies Intended to Manage Industry-Funded Bias." *Philosophy of Science* 85 (5): 869–81. <https://doi.org/10.1086/699713>.
- Hu, Jon. 2023. "Alzheimer's Drug Approvals Show We Need a Reevaluation of Patient Advocacy." *Stat*, December 18. <https://www.statnews.com/2023/12/18/patient-advocacy-alzheimers-lecanemab-aducanamab-fda/>.

- Hwang, Thomas J., Rachel E. Sachs, and Aaron S. Kesselheim. 2017. "Public Participation in Drafting of the 21st Century Cures Act." *Journal of Law Medicine & Ethics* 45 (2): 212–20. <https://doi.org/10.1177/1073110517720650>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Laursen, Bethany K., Chad Gonnerman, and Stephen J. Crowley. 2021. "Improving Philosophical Dialogue Interventions to Better Resolve Problematic Value Pluralism in Collaborative Environmental Science." *Studies in History and Philosophy of Science* 87: 54–71. <https://doi.org/10.1016/j.shpsa.2021.02.004>.
- Le Bihan, Soazig. 2024. "How to Not Secure Public Trust in Science: Representative Values versus Polarization and Marginalization." *Philosophy of Science* 91 (5): 1128–38. <https://doi.org/10.1017/psa.2023.119>.
- Liu, Ian T.T., Aaron S. Kesselheim, and Edward R. Scheffer Cliff. 2024. "Clinical Benefit and Regulatory Outcomes of Cancer Drugs Receiving Accelerated Approval." *JAMA* 331 (17): 1471–79. <https://doi.org/10.1001/jama.2024.2396>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Lynch, Holly Fernandez, and Alison Bateman-House. 2020. "Facilitating Both Evidence and Access: Improving FDA's Accelerated Approval and Expanded Access Pathways." *Journal of Law Medicine & Ethics* 48: 365–72. <https://doi.org/10.1177/1073110520935352>.
- Madl, Amy C. 2019. "Using Value-Agnostic Incentives to Promote Pharmaceutical Innovation." *Stanford Law Review* 71 (5): 1305–51.
- Mao, Xiangyun, G. Caleb Alexander, and Guanqiao Li. 2024. "Accelerated Approvals: Early-Phase Success or Premature Authorization?" *Cancer Cell* 42 (11): 1799–802. <https://doi.org/10.1016/j.ccell.2024.09.005>.
- Mostaghim, Sana R., Joshua J. Gagne, and Aaron S. Kesselheim. 2017. "Safety Related Label Changes for New Drugs after Approval in the US through Expedited Regulatory Pathways: Retrospective Cohort Study." *BMJ* 358: j3837. <https://doi.org/10.1136/bmj.j3837>.
- Naci, Huseyin, Olivier J. Wouters, Radhika Gupta, and John P.A. Ioannidis. 2017. "Timing and Characteristics of Cumulative Evidence Available on Novel Therapeutic Agents Receiving Food and Drug Administration Accelerated Approval." *The Milbank Quarterly* 95: 261–90. <https://doi.org/10.1111/1468-0009.12261>.
- Omae, Kenji, Akira Onishi, Ethan Sahker, and Toshi A. Furukawa. 2022. "US Food and Drug Administration Accelerated Approval Program for Nononcology Drug Indications between 1992 and 2018." *JAMA Network Open* 5 (9): e2230973. <https://doi.org/10.1001/jamanetworkopen.2022.30973>.
- Pielke, Roger A. 2007. *The Honest Broker: Making Sense of Science in Policy and Politics*. Cambridge University Press.
- Poongothai, Subramani, Ranjit Mohan Anjana, Ramasamy Aarth, et al. 2023. "Strategies for Participant Retention in Long Term Clinical Trials: A Participant-Centric Approaches." *Perspectives in Clinical Research* 14: 3–9. https://doi.org/10.4103/picr.picr_161_21.
- Porter, Laura D. 2022. "The Importance of Patient Engagement to Improve Healthcare Research and Safety." *Global Journal on Quality and Safety in Healthcare* 5: 27–30. <https://doi.org/10.36401/JQSH-22-X1>.
- Preziosi, Anthony J., and Ronny Priefer. 2024. "Oncology's Trial and Error: Analysis of the FDA Withdrawn Accelerated Approvals." *Life Sciences* 346: 122615. <https://doi.org/10.1016/j.lfs.2024.122615>.
- Rabinovici, Gil D. 2021. "Controversy and Progress in Alzheimer's Disease—FDA Approval of Aducanumab." *New England Journal of Medicine* 385: 771–74. <https://doi.org/10.1056/NEJMp2111320>.
- Rodriguez, Ryan, Rachel Brunner, Samantha Spencer, and Dima M. Qato. 2021. "Time to Inclusion in Clinical Guidance Documents for Non-Oncological Orphan Drugs and Biologics with Expedited FDA Designations: A Retrospective Survival Analysis." *BMJ Open* 11 (12): e057744. <https://doi.org/10.1136/bmjopen-2021-057744>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Scott, Emma C., Andrea C. Baines, Yutao Gong, et al. 2023. "Trends in the Approval of Cancer Therapies by the FDA in the Twenty-First Century." *Nature Reviews Drug Discovery* 22: 625–40. <https://doi.org/10.1038/s41573-023-00723-4>.

- Shahzad, Mahnum, Huseyin Naci, and Anita K. Wagner. 2021. “Estimated Medicare Spending on Cancer Drug Indications with a Confirmed Lack of Clinical Benefit after US Food and Drug Administration Accelerated Approval.” *JAMA Internal Medicine* 181: 1673–75. <https://doi.org/10.1001/jamainternmed.2021.5989>.
- Skydel, Joshua J., Alexander C. Egilman, Joshua D. Wallach, Reshma Ramachandran, Ravi Gupta, and Joseph S. Ross. 2022. “Spending by the Centers for Medicare and Medicaid Services before and after Confirmation of Benefit for Drugs Granted US Food and Drug Administration Accelerated Approval, 2012 to 2017.” *JAMA Health Forum* 33 (5): e221158. <https://doi.org/10.1001/jamahealthforum.2022.1158>.
- Tempini, Niccolo, and David Teira. 2019. “Is the Genie Out of the Bottle? Digital Platforms and the Future of Clinical Trials.” *Economy and Society* 48 (1): 77–106. <https://doi.org/10.1080/03085147.2018.1547496>.
- Torreele, Els. 2024. “Why Are Our Medicines So Expensive? Spoiler: Not for the Reasons You Are Being Told....” *European Journal of General Practice* 30 (1): 2308006. <https://doi.org/10.1080/13814788.2024.2308006>.
- Wedam, Suparna, Lola Fashoyin-Aje, Erik Bloomquist, et al. 2020. “FDA Approval Summary: Palbociclib for Male Patients with Metastatic Breast Cancer.” *Clinical Cancer Research* 26 (6): 1208–12. <https://doi.org/10.1158/1078-0432.CCR-19-2580>.
- Wehling, Peter, Willy Viehöver, and Sophia Koenen. 2015. *The Public Shaping of Medical Research: Patient Associations, Health Movements and Biomedicine*. Taylor & Francis Group.
- Will, Catherine, and Tiago Moreira. 2010. *Medical Proofs, Social Experiments: Clinical Trials in Shifting Contexts*. Ashgate.
- Woloshin, Steven, Lisa M. Schwartz, Brian White, and Thomas J. Moore. 2017. “The Fate of FDA Postapproval Studies.” *New England Journal of Medicine* 377: 1114–17. <https://doi.org/10.1056/NEJMp1705800>.

20

SCIENCE AND DEMOCRACY

Hannah Hilligardt and Torsten Wilholt

20.1 Introduction

When considering how values manifest themselves in science, the interplay between scientific and democratic institutions and processes plays a special role. The aim of this chapter is to clarify what this interplay entails in order to develop an account of how science should be organized that is explicitly grounded in political theory (rather than ethics or epistemology). To do so, we will proceed in the following way: In the next section, we will explore the potential science and democracy have to support each other's values. We quickly leave this perspective on their relation behind to take a more functionalist approach starting with [Section 20.3](#): Science provides information that is needed as a basis for democratic decision-making and thus has an important role to fulfill in democracy. In [Section 20.4](#), we will derive normative consequences of this role of science, which include the independence of science, its responsiveness to citizens' values, preferences, and interests, and its commitment to an orientation toward justice. As is customary in the philosophy of science, we first develop these points in accordance with a liberal understanding of democracy. In [Section 20.5](#), however, we also point out the limitations inherent in this prevailing view and suggest some benefits for a philosophy of science that also takes alternatives to traditional liberal political philosophy, such as non-ideal, communitarian, or agonistic approaches, seriously as philosophical resources.

20.2 Science and democracy: Ideas about their kinship

A close affinity and a relationship of mutual dependency have often been asserted between science and democracy. Strong claims in this vein were emphasized in the middle of the last century, often in the spirit of defending science and democracy alike from the threats of totalitarian systems. For example, American sociologist Bernard [Barber \(1952\)](#) suggested that science and the democratic political order go hand in hand, based on the assertion that both share core values such as individualism, rationality, universalism, and progress. Robert [Merton's \(1938, 327\)](#) earlier and more famous reflections that non-democratic systems in their influence on the sciences must inevitably come into conflict with the institutionalized norms of science and its ethos also fit into this general spirit. At many different points in the history of liberal thought, one finds authors who combine the claim of a

close association between science and democracy with an enthusiastic partisanship for both—such as when Thomas Jefferson describes freedom itself as the “first-born daughter of science” or when a handbook of the liberal democratic movement in Germany declares in 1848 that science could only have originated in ancient Greece, since only there the necessary liberal political conditions had obtained (Jefferson 1896, 3; Scheidler 1848, 622).

If the idea of a connection between science and democracy is understood simply as the assertion of a straightforward conditional relationship (in one direction or the other), then it is clearly not credible. Writing at around the same time as Merton, John Dewey had already declared, “It is no longer possible to hold the simple faith of the Enlightenment that assured advance of science will produce free institutions [...]” (Dewey 1988, 156). Dewey drew attention to the technical applications of science and their potential to cause great harm to free institutions, for example by bringing about a high concentration of capital or by providing the world’s dictators with new, powerful propaganda tools. The conditional in the other direction, i.e., the assertion that democracy is necessary for science, is just as unconvincing. After all, even the celebrated achievements of the Scientific Revolution were accomplished in repressive social and political environments, with liberal democracy nowhere in sight, and the more recent past has only discredited the thesis more. If you think of scientific efforts to develop weapons of mass destruction in National Socialist Germany and other undemocratic systems, what is most frightening about them is not how badly, but how well scientific research can function in totalitarian regimes (cf. Sarewitz 2003, 186–7).

To be convincing today, the claimed connection between science and democracy must therefore be read in a weaker sense. Dewey himself had some noteworthy thoughts on this. According to him, the two are based on common epistemic principles, which he summarizes in a characterization of the “scientific attitude.” For him, this explicitly includes the

willingness to hold belief in suspense, ability to doubt until evidence is obtained; willingness to go where evidence points instead of putting first a personally preferred conclusion; ability to hold ideas in solution and use them as hypotheses to be tested instead of as dogmas to be asserted; and (possibly the most distinctive of all) enjoyment of new fields for inquiry and of new problems.

(Dewey 1988, 166)

Dewey assumes that a functioning democracy requires, among other things, that citizens are sufficiently and independently informed. Since he believes that the necessary knowledge of citizens can only flourish if the epistemic attitude described above is widely disseminated and recognized, democracy is dependent on the scientific attitude. In a discussion that bears similarities to that of Dewey, Bernard Williams attributes to the sciences a particularly striking manifestation of the “virtues of truth” (Williams 2002, 144). The preoccupation with science confronts a subject in a particularly striking way with a situation in which truth is not at the disposal of any personal counterpart. Science, Williams suggests, is a practice in which that particular epistemological stance can be learned in an exceptionally straightforward way.

The claim that science can make constructive contributions to a democratic culture in this way should of course not be confused with the naïve idea that mastery of science makes one immune to anti-democratic and authoritarian tendencies. The history of science is rich in depressing counterexamples (cf. Beyerchen 1977; Proctor 1988). Nor is it possible to draw a clear and simple conclusion as to how much the potentially beneficial contributions of the sciences to epistemic culture weigh against the threat to democracy and human life posed by the technological fruits of science that Dewey had so presciently addressed.

20.3 Science's role in liberal democracies

This is the reason why, in the remainder of this chapter, we will take as our starting point an even weaker link between science and democracy and focus on political philosophers, especially in the liberal tradition, who have considered the relationship between science and democracy primarily in terms of the *function* or *role* that science has in a democracy: To provide knowledge. Let us briefly elaborate on why knowledge is so important from a democratic perspective; in the subsequent section we will see what implications this political role has for science.

Scientific knowledge is needed to improve the ability of citizens to assert their values and interests within the democratic process. If the preferences that citizens form and bring to bear in the political process are supposed to reflect their values and interests, then they must be well-informed, that is, they must be based on reliable belief (including, at the very least, reliable views on how government actions or legislative decisions are likely to affect the lives of citizens; cf. [Fuerstein 2008](#), 76; [Goldman 1999](#), 320–5). As the consequences of political decisions are generally causally delivered via complex natural, technological, social, and economic processes, the availability of *scientific* knowledge specifically belongs to the prerequisites for the well-informedness of citizens in modern societies. The importance of knowledge is particularly accentuated in deliberative theories of democracy, which see the political process as public use of reason and an exchange of arguments and consider the formation of citizens' preferences to be an essential part of the democratic process. But even in more parsimonious conceptions of democracy, which focus on voting and elections and regard the democratic process as a means of aggregating existing preferences, citizens' access to knowledge is emphasized as a prerequisite (e.g. [Dahl 1989](#), 112).

The general approach to describing the connection between science and democracy in terms of the relevance of scientific knowledge for the democratic process is embedded in a liberal understanding of the state and politics in the broadest sense. By this we mean the endeavor to develop the justification of state power and the answering of normative questions about political relations and institutions without relying on a comprehensive understanding of the good (so as not to commit the citizens to such an understanding). This core approach characterizes the liberal tradition from the beginnings of Hobbes' and Locke's reflections on limited government and constitutionalism to more recent attempts, such as those of Rawls and Habermas, to in one way or another find normative foundations of political philosophy in what citizens must minimally agree upon (cf. [Heath 2022](#), chap. 3).

Of course, science has other important potentials besides its role for the democratic process: It can inform not only collective and political decisions but also individual ones; it can contribute to collective prosperity and the improvement of health and living conditions; and it can promote our understanding of the world and our own position in it. But all these services (as well as the harms and threats it poses) are provided as a collective endeavor shaped and enabled by institutions without which contemporary science would not be possible. These include, above all, many informal and evolved institutions, but also formal ones in some key places, whose rules can ultimately only be enforced by the state ([Mantzavinos 2024](#)). In democratic states, the collective enterprise of science as a whole is in this sense subject to the democratic order that makes it possible in the first place.

The idea that citizens of a democracy are dependent on scientific knowledge in order to be able to effectively assert their interests does not amount to a glorification of science as the royal road to democracy, nor does it presuppose that science is politically neutral or value-free. It is not assumed that scientific knowledge puts citizens in a perfect epistemic position, but only that they are relevantly better off epistemically with access to scientific knowledge than without it. For this to be the case, the science in question must fulfill certain requirements, which will interest us in detail in the next section.

20.4 Implications for philosophy of science

Clarifying the role science occupies in and for democratic systems has been productive for philosophers of science because a number of ideals concerning science's institutional structure, scientific norms, and value management can be derived from this role. In this way, answers to questions that play a central role in the debates on science and values and on trust in science can be found that have their basis in political theory (rather than ethics or epistemology). The main line of thought that we will follow in this chapter is therefore that the institutional structures, the norms and rules that make up the institutions of science, and the way in which values and interests are dealt with in internal scientific decisions ought to be determined by the functions served by science for a just democratic polity.

First of all, considerations about science's political role are crucial in order to specify its proper relation to (other) political institutions. If the knowledge that citizens obtain from scientific sources is to enable them to bring their values and interests to bear in the democratic process, then the production and dissemination of this knowledge must be independent of the political powers. In modern societies, where the results of scientific research are often decisive for the assessment of controversial issues, there are strong incentives for political forces to use their power to influence what science does. Since the inception of modern democratic institutions, the protection of science from political interference has therefore been called for—e.g. by Thomas Jefferson, James Madison, and others in America, and in France by Condorcet, who was commissioned by the French National Assembly to make proposals for the institutional foundations for science and education (Wilholt 2010, 2012). The idea was also taken up by German liberal thinkers such as Julius Fröbel, who declared in 1846, “The combination of these two principles, the rule of majorities in the realm of ends, and the rule of freedom and anarchy in the realm of theory, constitutes the only correct method of political life and progress” (Fröbel 1850, 107). Following the 1848 revolution, the constitutional assembly in Frankfurt's Paulskirche, which included Fröbel and many other liberals, included the principle that “Science and its teaching is free” on its list of fundamental rights. Although the Paulskirche constitution never came into force, its catalog of basic rights influenced later constitutions. Since then, academic freedom has been constitutionally recognized in many continental European countries.

It is worth pointing out that this tradition of justifying the need for freedom and independence of science on democratic grounds is different from epistemological arguments that usually dominate Anglophone discussions about freedom and independence in science. The latter typically start by emphasizing the fallibility of inquiry (often with reference to John Stuart Mill) and the unpredictability of what might later become of a nascent idea or line of research. Building on this, the arguments advise that any epistemic collective enterprise would do best not to suppress any approach or idea altogether, even if it seems false, implausible, or unpromising at the moment, but instead to hedge its bets, achieving the required diversity through individual freedom, as this unleashes creative potential and allows individuals to make the most of their local knowledge. Epistemological arguments of this kind have become the dominant narrative in support of the independence of science in the post-war period—following Vannevar Bush's successful promotion of a “New Social Contract for Science” (Guston 2000, chap. 2), Michael Polanyi's fight against J. D. Bernal's plans for socialization and greater planning of scientific research (Fehér 1996), and other Cold War developments.

In contrast, the argumentation outlined above following Condorcet, the American Founding Fathers, the German liberals of the nineteenth century, and others, which can be described as a political argument for the independence of science (Wilholt 2010, 2012), amounts to demanding a separation of powers between science and the (other) political powers (cf. Brown and Guston 2009). While epistemic arguments emphasize individual creativity and the exploitation of individual knowledge and thus call for individual freedoms, an argument based on democratic theory strengthens a

corporative independence of science. It must be free from the kind of influence by political powers and similarly powerful actors that would allow it to be steered toward certain outcomes. This strengthens the demand for freedom in the choice of research questions and for the public nature of the results of scientific research. It may also be taken to strengthen the case for pluralism within the scientific community, in so far as pluralism provides good grounds for optimism that scientific information effectively improves the politically relevant knowledge base of citizens. This is so because admitting a plurality of different approaches both facilitates mechanisms of mutual criticism (and thus improves epistemic quality) and ensures that even in a society with very different value outlooks, as many people as possible have access to the epistemic resources they need to participate effectively in deliberation (Hilligardt 2023; Longino 2002, chap. 6; Rolin 2021; Thoma 2024; Wilholt 2012, 247–52).

A second type of normative consequence that can be derived from the role of science for democracy concerns the responsibilities and duties that scientists have toward the public. They can differ, as S. Andrew Schroeder (2020) has pointed out, from moral responsibilities that apply to all moral agents. Thus, science as an institution arguably has the duty to be responsive to the public's preferences and values. Kitcher famously and influentially spelt out how this can be done in his ideal of “well-ordered science.” According to this ideal (Kitcher 2001), the correct answer to the question of which epistemic objectives are classified as significant and thus made the target of scientific research is the one that a collective of idealized, well-informed deliberators would agree on at the end of a discussion in which different points of view have been expressed and weighed against each other. In Kitcher 2011, the same is also extended to the question of what the appropriate standards of evidence are by which scientific outcomes are determined, and questions about which knowledge is applied and in what way. For Kitcher, these are first and foremost purely theoretical scenarios that are intended to show in what sense it is even possible to speak of answers to the above questions that are distinguished as correct. When it comes to approaching this ideal in the real world, Kitcher suggests relying on participatory approaches. In particular, he discusses the idea of creating “groups of citizen representatives, drawn from diverse segments of different societies,” that are subjected to a specific instructional process in which they form a basic understanding of relevant research processes, and who are allowed to go “behind the scenes” of research (Kitcher 2011, 129). These special efforts are to ensure that science is not subjected to a tyranny of the ignorant. Others, too, have made proposals for how to ensure science's responsiveness to public values, ranging from stakeholder participation procedures, the involvement of randomly selected representative samples of the population to the consultation of social science research on public values (cf. Bozeman and Sarewitz 2011; Intemann 2015; Lusk 2021; Rolin 2021; Schroeder 2021).

A different line of argument emphasizes science's duty to serve the public in a way that takes into account requirements for fairness and equal treatment. At a very fundamental level, equality (equality of opportunity for political participation, equality before the law) is one of the basic building blocks of democracy according to the liberal understanding. For this reason alone the role of science in democracy must also be considered in light of these principles (Kitcher 2011, chap. 3). Liberal theorists have also argued that the sense of equality relevant to democracy must also include the fulfillment of a minimum level of *social* equity, because certain material requirements must be met in order to achieve equal opportunity of political participation (Bohman 1997; Waldron 1993).

And even more specifically, it is compelling that equitable access to *epistemic* resources is required (Thoma 2024). The “epistemic basic structure,” Irzik and Kurtulmus argue, “should serve individuals fairly and provide them with an equal opportunity to gain knowledge in order to reason about the common good, their individual good and pursuit thereof” (2024, 4). Specifying the ideal distribution of epistemic resources (Irizik and Kurtulmus draw on a Rawlsian theory of distributive justice) can help identify and criticize unequal and unfair distributions. Kitcher (2001, 128) similarly

highlights “the problem of inadequate representation,” such as the disproportionate funding for mild diseases affecting a small number of solvent people (e.g. acne) versus deadly diseases like malaria and tuberculosis impacting larger, less affluent populations (Kitcher 2011; Reiss and Kitcher 2010).

The justice-based approach has also been used to evaluate the legitimacy of values in “internal” phases of research such as the management of epistemic risks. Ahmad Elabbar (forthcoming), drawing on Irzik and Kurtulmus (2024), argues that justice considerations can justify varying the evidential thresholds set by the Intergovernmental Panel on Climate Change for inclusion of hypotheses or projections in their reports. Elabbar demonstrates that the pronounced inequality in climate data and research between the Global North and South means the IPCC harms the fundamental interests of Global South citizens by applying uniform evidence standards across regions.

In sum, based on the importance of scientific knowledge for effective democratic self-governance, a strong case can be made that science ought to be organized in a way that allows for it to produce independent knowledge that meets the population’s interests and addresses justice issues. Of course, achieving this ideal is far from a simple task in societies with deep inequalities in power and opportunity, significant value differences, and widespread opinion manipulation and misinformation. We have offered some ideas in this section—a comprehensive account of the way in which science ought to be organized in a democracy goes beyond the scope of this chapter. Before coming to an end, however, we want to offer some general reflections on the tools that philosophers employ when engaging with the question of how science can best fulfill its role.

20.5 De-idealizing political philosophy of science

Philosophers of science rarely discuss the political philosophy resources themselves and the limitations that might arise from them. Additionally, it is often overlooked how much the perspective is informed by assumptions about the nature of society and politics—a problem that is particularly visible in conflicts between philosophers of science and science and technology scholars. To illustrate this, we will use the rest of this chapter to discuss two prominent critiques of liberal democratic theory and their implications for philosophy of science views on science and democracy.

Liberal theories of democracy, in particular John Rawls’ and Jürgen Habermas’ versions, which so often find application in philosophy of science, have been debated for decades. One key point of contention is their use of idealizations. Charles Mills (2005) criticizes that Rawls does not merely formulate an ideal for society in the sense of a desired state of affairs but draws up an *idealized model of society*. Rawls envisions deliberators to be perfectly rational and disinterested and justifies principles of justice on the grounds of the imagined outcome of their deliberations. Mills argues that these idealizations are “abstracting away from realities crucial to our comprehension of the actual workings of injustice in human interactions and social institutions” (Mills 2005, 170), making them ineffective for addressing real-world injustices.

Most philosophers of science are aware of the debate between ideal and non-ideal theory and those who draw on Rawls’ work often acknowledge this in footnotes (see, for example, Cabrera 2022, 816; Irzik and Kurtulmus 2024, 4). However, it is well worth considering non-ideal theorists’ critique in more detail than footnotes permit, as it can have relevant consequences for the characteristics of what we consider well-ordered science. The question is how much the normative guidance from philosophers of science relies on idealized assumptions about scientists and society. Do these idealizations limit the potential to address real-world problems? A key example which highlights the importance of these considerations is the idealization of autonomy.

A common characteristic of many liberal theories of democracy is to imagine individuals as autonomous and independent agents, or to at least posit this as an ideal to strive toward. The so-called communitarian critique (cf. Mills 2005; Sandel 1982) therefore condemned “much liberal political

theory for conceptualizing the individual person as prior to and not fundamentally conditioned by commitments and values of particular social groups,” as Iris Marion Young (1995, 182) summarizes.

This individualism, arguably, has also carried into philosophy of science, firstly in the way researchers used to be idealized. Feminist scholars have famously argued that research is shaped in crucial ways by the social situatedness of the researchers and have grounded this view explicitly in non-ideal political theories, in particular Marxist theories of the self (see Harding 1992; Hartsock 1983). They have used this argument to critique the notion that science is value-free and by doing so have contributed substantially to the values in science debate we have today.

More importantly for the role of science in democracies, however, it has been argued that the idealization of *citizens* as autonomous agents is also reflected in the way many philosophers of science imagine lay people, especially in the context of scientific advice and science communication. Elabbar (2023, 2) suggests that much of the existing literature on scientific advice “share[s] a common commitment to the principle of decisional autonomy: advisors should resolve value-laden choices in a manner that preserves the decisional autonomy, or self-determination, of advisees,” for instance by making their own values transparent. This commitment to autonomy has been questioned from a non-ideal perspective (see especially McKenna 2023). For, as empirical research suggests, individuals do not form preferences and judgments independently and autonomously, nor do they incorporate scientific information into their preferences in a context-independent manner. Group affiliations and loyalties—for instance to class, race, and religion—are key factors determining lay people’s party affiliations, which in turn often determine attitudes toward scientific information (Achen and Bartels 2017; McKenna 2023; Williams 2023). In the US, for instance, almost all “liberal Democrats” believe that global warming is real, but less than half of all “conservative Republicans” do (cf. McKenna 2023, 66). This group-based thinking also occurs among those who are highly politically engaged and informed, i.e., those who come closest to the ideal, well-informed citizen (Williams 2023, 3).

One conclusion that has been derived from this is that scientists have to take a more pro-active stance in advising and informing public and private decision-making. Norms such as neutrality and non-partisanship, as well as measures aimed at increasing transparency or using public participation to resolve value disputes in research, do not necessarily lead to autonomous and independent decision-making (Elabbar 2023), but rather to decision-making that conforms with existing beliefs and societal power relations (Hilligardt 2023; Rolin 2021). Scientific information that conflicts with widely held beliefs and vested interests often struggles to be taken seriously. Consequently, methods of persuasion and marketing, as well as outspoken advocacy, may be necessary to ensure that relevant scientific information is not only available but also incorporated into decision-making (McKenna 2023; Oppenheimer et al. 2019).¹

This critique highlights that philosophical frameworks for well-ordered science must strike a delicate balance between offering normative guidance and avoiding unrealistic idealizations. While there are valid reasons to view citizens as autonomous decision-makers and to model science based on this view, it should not be assumed that this model is always the most suitable (see also Biddle 2009). Political philosophy, with its long history of critically examining various models of society, can offer valuable insights for reflection in this area.

A second non-ideal critique of liberal theories, relevant to science and democracy, addresses how consensus and conflict are managed in democratic decision-making. This so-called agonistic critique has been influential in science and technology studies, for instance in the more recent work of Bruno Latour (2018). Based on this perspective not only the characterization of well-ordered science may be put into question but the very role that is attributed to science in the context of liberal democratic accounts.

Agonistic thinkers such as Chantal Mouffe (2000, 2005) argue that conflict is essential for forming and sustaining personal and group identities. Conflict shouldn't be seen as inherently negative or something that must be resolved. Of course, liberal theories, too, allow for disagreement—indeed value pluralism is one of the core tenets of liberalism. But they draw a line between issues open to reasonable disagreement (such as values and religion) and those that every rational person should agree on. Agonistic theorists find this distinction untenable and covertly exclusionary. They argue that those who don't accept what is deemed reasonable are discredited as irrational and excluded from the debate. Mouffe (2000, 756) states that while pluralist democracy requires some consensus on ethical-political principles, these principles are always subject to conflicting interpretations and thus agreement on them can only come in the form of a “conflictual consensus.” Therefore, pluralist democracy must accommodate dissent and the institutions that facilitate its expression. In essence, Mouffe contends that nothing should be considered beyond political contestation; we can only temporarily agree not to debate certain issues (Mouffe 2000, 756; see also Schönwitz 2022 for an application of this argument in the context of the values in science debate).

This is relevant to the role ascribed to science because, at least at times, philosophers of science argue that scientific information should ideally not be subject to political debate but instead form the common ground that political debate starts from. Stephen John (2021a, 5), for example, argues that some “factual claims” should be “beyond political contestation” for the sake of fruitful political debate: “For example, even if value-based disagreements might lead to disagreements over climate policy, such disagreements require some epistemic common ground, say, that climate change is occurring at all.” The agonistic perspective casts doubt on the notion that science should even aspire to play such a role if it presupposes exclusion from the realm of political contestation.

It is debatable whether it is wise to remove any science-related discourse boundaries in a world where strategic misinformation is so ubiquitous (see Biddle and Leuschner 2015; de Melo-Martín and Intemann 2014). Nevertheless, using rationality and expertise as common ground in political debates can clearly be exclusionary, especially when some groups in society are significantly better trained to prevail in debates on these terms than others (cf. Young 2000, chap. 1) and when scientific information itself is at times biased in favor of specific interests (e.g. Stegenga 2017, chap. 10; Wilholt 2009 chap. 10). Agonistic theories highlight these challenges; even proponents of scientific discourse boundaries must consider them to prevent unjust exclusion in the name of science.

20.6 Conclusion

We started by reviewing some sweeping claims about the special kinship between science and democracy that have historically been made, to then identify Dewey's thesis of a common epistemic attitude, which is conducive to both democracy and science, as the most plausible, if still somewhat vague, core of this idea. We then moved on to focus on a connection between science and democracy that can be described and substantiated much more precisely: Namely, that science plays important epistemic roles within democracy. We have shown that this provides a particular justification for the need for the political independence of science, as well as a derivation of the duties of scientists to be responsive to the needs and interests of the public, and to bring considerations of justice to bear in the production of knowledge. These things can be well explained, especially against the background of a liberal understanding of democracy.

But liberal theories of democracy have been subject of contestation for many decades in political philosophy. In light of these critiques, we have argued that philosophers of science who comment on science's role in democracies need to reflect on assumptions, abstractions, and idealizations they adopt from the political resources they use. Concretely, we have considered critiques

of two idealizations that are prominent in theories of liberal democracy as well as in philosophy of science: A focus on and appraisal of autonomous individuals and the aspiration of establishing a non-political common ground or consensus that limits the range of legitimate conflict. These critiques are meant to demonstrate the extent to which both philosophers' account of well-ordered science and the very role that is attributed to science hinges on what is considered a good democratic system and thereby support the view that philosophers of science should pay attention to political philosophy.

Note

- 1 It might be objected that respecting others' autonomy is an ethical principle that ought to be upheld independent of the consequences of doing so. Stephen John (2021b) has argued along these lines in a paper on scientific deceit. When looking at individual instances, such argument remains forceful. Many non-ideal approaches, however, prioritize the societal perspective. Upholding a certain principle becomes problematic when it systematically disadvantages or brings about negative consequences for part of society. As Mills (2005, 172) argues, philosophers ought to ask "cui bono?" when assessing an idealization and ideal theory more broadly.

References

- Achen, Christopher, and Larry Bartels. 2017. *Democracy for Realists*. Princeton University Press.
- Barber, Bernard. 1952. *Science and the Social Order*. Collier.
- Beyerchen, Alan D. 1977. *Scientists Under Hitler: Politics and the Physics Community in the Third Reich*. Yale University Press.
- Biddle, Justin B. 2009. "Advocates or Unencumbered Selves? On the Role of Mill's Political Liberalism in Longino's Contextual Empiricism." *Philosophy of Science* 76 (5): 612–23. <https://doi.org/10.1086/605791>.
- Biddle, Justin B., and Anna Leuschner. 2015. "Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?" *European Journal for Philosophy of Science* 5 (3): 261–78. <https://doi.org/10.1007/s13194-014-0101-x>.
- Bohman, James. 1997. "Deliberative Democracy and Effective Social Freedom: Capabilities, Resources, and Opportunities." In *Deliberative Democracy: Essays on Reason and Politics*, edited by James Bohman and William Rehg. MIT Press.
- Bozeman, Barry, and Daniel Sarewitz. 2011. "Public Value Mapping and Science Policy Evaluation." *Minerva* 49 (1): 1–23. <https://doi.org/10.1007/s11024-011-9161-7>.
- Brown, Mark B., and David H. Guston. 2009. "Science, Democracy, and the Right to Research." *Science and Engineering Ethics* 15: 351–66. <https://doi.org/10.1007/s11948-009-9135-4>.
- Cabrera, Frank. 2022. "A Rawlsian Solution to the New Demarcation Problem." *Canadian Journal of Philosophy* 52 (8): 810–27. <https://doi.org/10.1017/can.2023.19>.
- Dahl, Robert A. 1989. *Democracy and Its Critics*. Yale University Press.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2014. "Who's Afraid of Dissent? Addressing Concerns About Undermining Scientific Consensus in Public Policy Developments." *Perspectives on Science* 22 (4): 593–615. https://doi.org/10.1162/POSC_a_00151.
- Dewey, John. 1988. "Freedom and Culture." In *The Later Works, 1925–1953, Volume 13, 1938–1939*, edited by Jo A. Boydston. Southern Illinois University Press.
- Elabbar, Ahmad. forthcoming. "Varying Evidential Standards as a Matter of Justice." *The British Journal for the Philosophy of Science*. <https://doi.org/10.1086/727429>.
- Elabbar, Ahmad. 2023. "The Curatorial View of Assessment and the Ethics of Scientific Advice: Beyond Decisional Autonomy towards Distributive Epistemic Justice." iuhpst.org/media/pdf/Elabbar_IUHPST_Prize_2023.pdf.
- Fehér, Márta. 1996. "Science and Liberalism: Michael Polanyi on the Freedom of Science." *Polanyiana* 5 (1): 47–62.
- Fröbel, Julius. 1850. *System der sozialen Politik* 2. Verlagsbureau.
- Fuerstein, Michael. 2008. "Epistemic Democracy and the Social Character of Knowledge." *Episteme* 5 (1): 74–93. <https://doi.org/10.3366/E1742360008000245>.
- Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford University Press.

- Guston, David H. 2000. *Between Politics and Science: Assuring the Integrity and Productivity of Research*. Cambridge University Press.
- Harding, Sandra G. 1992. "After the Neutrality Ideal: Science, Politics, and 'Strong Objectivity'." *Social Research* 59 (3): 567–87. <https://www.jstor.org/stable/40970706>.
- Hartsock, Nancy. 1983. "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism." In *Discovering Reality: Feminist Perspectives on Epistemology, Metaphysics, Methodology, and Philosophy of Science*, edited by Merrill B. Hintikka and Sandra G. Harding. D. Reidel.
- Heath, Joseph. 2022. *The Machinery of Government: Public Administration and the Liberal State*. Oxford University Press.
- Hillgardt, Hannah. 2023. "Partisan Science and the Democratic Legitimacy Ideal." *Synthese* 202 (5). <https://doi.org/10.1007/s11229-023-04370-5>.
- Intemann, Kristen. 2015. "Distinguishing Between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Irzik, Gürol, and Faik Kurtulmus. 2024. "Distributive Epistemic Justice in Science." *The British Journal for the Philosophy of Science* 75 (2). <https://doi.org/10.1086/715351>.
- Jefferson, Thomas. 1896. *The Writings of Thomas Jefferson*. 7. Edited by Paul L. Ford. G. P. Putnam's Sons.
- John, Stephen. 2021a. "Science, Politics and Regulation: The Trust-Based Approach to the Demarcation Problem." *Studies in History and Philosophy of Science* 90: 1–9. <https://doi.org/10.1016/j.shpsa.2021.08.006>.
- John, Stephen. 2021b. "Scientific Deceit." *Synthese* 198 (1): 373–94. <https://doi.org/10.1007/s11229-018-02017-4>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus; Roundhouse.
- Latour, Bruno. 2018. *Facing Gaia: Eight Lectures on the New Climatic Regime*. John Wiley & Sons.
- Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Mantzavinos, Chrysostomos. 2024. *The Constitution of Science*. Cambridge University Press.
- McKenna, Robin. 2023. *Non-Ideal Epistemology*. Oxford University Press.
- Merton, Robert K. 1938. "Science and the Social Order." *Philosophy of Science* 5 (3): 321–37. <https://doi.org/10.1086/286513>.
- Mills, Charles W. 2005. "Ideal Theory' as Ideology." *Hypatia* 20 (3): 165–83. <https://doi.org/10.1111/j.1527-2001.2005.tb00493.x>.
- Mouffe, Chantal. 2000. "Deliberative Democracy or Agonistic Pluralism." *Social Research* 66 (3): 745–758. <http://www.jstor.org/stable/40971349>.
- Mouffe, Chantal. 2005. *On the Political*. Routledge. <https://doi.org/10.4324/9780203870112>.
- Oppenheimer, Michael, Naomi Oreskes, Dale Jamieson, et al. 2019. *Discerning Experts: The Practices of Scientific Assessment for Environmental Policy*. The University of Chicago Press.
- Proctor, Robert N. 1988. *Racial Hygiene: Medicine Under the Nazis*. Harvard University Press.
- Reiss, Julian, and Philip Kitcher. 2010. "Biomedical Research, Neglected Diseases, and Well-Ordered Science." *Theoria* 24 (3): 263–82. <https://doi.org/10.1387/theoria.696>.
- Rolin, Kristina H. 2021. "Objectivity, Trust and Social Responsibility." *Synthese* 199 (1–2): 513–33. <https://doi.org/10.1007/s11229-020-02669-1>.
- Sandel, Michael J. 1982. *Liberalism and the Limits of Justice*. Cambridge University Press.
- Sarewitz, Daniel. 2003. "Science and Happiness." In *Living with the Genie: Essays on Technology and the Quest of Human Mastery*, edited by Alan Lightman, Daniel Sarewitz, and Christina Desser. Island Press.
- Scheidler, Karl Hermann. 1848. "Universitäten." In *Staats-Lexikon*, edited by Carl von Rotteck and Carl Welcker. Vol. 12. Hammerich.
- Schönwittz, Tobias. 2022. "Values in Science, Biodiversity Research, and the Problem of Particularity." *Kennedy Institute of Ethics Journal* 32 (1): 69–101. <https://doi.org/10.1353/ken.2022.0003>.
- Schroeder, S. Andrew. 2020. "Thinking about Values in Science: Ethical Versus Political Approaches." *Canadian Journal of Philosophy*: 1–10. <https://doi.org/10.1017/can.2020.41>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Stegenga, Jacob. 2017. *Medical Nihilism*. Oxford University Press.
- Thoma, Johanna. 2024. "Social Science, Policy and Democracy." *Philosophy and Public Affairs* 52 (1): 5–41. <https://doi.org/10.1111/papa.12250>.

- Waldron, Jeremy. 1993. "Social Citizenship and the Defense of Welfare Provision." In *Liberal Rights: Collected Papers 1981–1991*, edited by Jeremy Waldron. Cambridge University Press.
- Wilholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science Part A* 40 (1): 92–101. <https://doi.org/10.1016/j.shpsa.2008.12.005>.
- Wilholt, Torsten. 2010. "Scientific Freedom: Its Grounds and Their Limitations." *Studies in History and Philosophy of Science* 41 (2): 174–81. <https://doi.org/10.1016/j.shpsa.2010.03.003>.
- Wilholt, Torsten. 2012. *Die Freiheit der Forschung: Begründungen und Begrenzungen*. Suhrkamp.
- Williams, Bernard. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton University Press.
- Williams, Daniel. 2023. "The Case for Partisan Motivated Reasoning." *Synthese* 202 (3). <https://doi.org/10.1007/s11229-023-04223-1>.
- Young, Iris Marion. 1995. "Rawls's Political Liberalism." *Journal of Political Philosophy* 3 (2): 181–90. <https://doi.org/10.1111/j.1467-9760.1995.tb00034.x>.
- Young, Iris Marion. 2000. *Inclusion and Democracy*. Oxford University Press.

21

SCIENCE AND RELIGION IN A DEMOCRATIC SOCIETY

Yiftach Fehige

21.1 Introduction

In what follows, I challenge the widely held belief that modern science—encompassing fields such as quantum physics, evolutionary biology, and polymer chemistry—fundamentally promotes democracy, whereas religious traditions—such as Judaism, Christianity, and Islam—necessarily pose a threat to it (for defenses of this view, see [Bronowski 1972](#); [Russell 1932, 1939](#); [Kitcher 2011](#)).

To this end, [Section 21.2](#) presents the theory that democracy is rooted in emancipative values, such as individual autonomy, equality, and freedom of choice. According to this view, the spread of emancipative values displaces autocratic values, such as respect for authority, hierarchy, and tradition. In this framework, modern science is not only defined by emancipative values but also serves as a driving force for their spread. Religion, by contrast, is characterized by autocratic values and is responsible for their persistence. On this account, science naturally aligns with democracy, whereas religion does not.

In [Section 21.3](#), I challenge this theory, particularly its characterization of science and religion. I argue that democracy is sustained by value judgments, which I define broadly as deliberative decisions that are informed by values and are often meant to resolve value conflicts. This means that the mere spread of emancipative values in a society does not, in itself, anchor democratic will formation. Moreover, the value judgments that constitute democratic will formation are shaped by both emancipative and autocratic values. Additionally, science is not exclusively defined by emancipative values. If these claims hold, then the stark dichotomy between science and religion in terms of their democratic potential becomes difficult to maintain.

In [Section 21.4](#), I offer reasons why the liberal state has good grounds to cultivate existing religious traditions for the sake of democracy. I will define the liberal state and explain how its defining elements support the inclusion of religious traditions in deliberations aimed at democratic will formation.

21.2 Democracy is rooted in emancipative values

It is probably impossible to overstate the political significance of the encounter between science and religion. [Lafont \(2022\)](#) rightly notes that a key issue in “debates about the proper place of religion in democratic societies” is “whether democracy and secularism are necessarily connected” (193).

Modern science enters the analysis of this issue because it is often perceived as the driving force behind secularization. In the words of [Ecklund and Long \(2011\)](#), scholars and public intellectuals almost uniformly perceive scientists as “the carriers of a secularist impulse” (254). Consequently, the clash between modern science and religious traditions becomes a most significant issue for democracy. As [Lafont \(2022\)](#) reflects, “if there is no hope that secular and religious citizens can take ownership of and identify with democratic institutions in equal measure, then the future of democracy within pluralist societies is seriously threatened” (194).

[Welzel \(2021\)](#) begs to differ, and he argues that in “mainstream scholarship” on democracy the dominance of pessimists have misdirected the discussion. According to him, pessimists are misled by the rise of “illiberal populism” in democracies like Germany and the United States, and the persistence of powerful autocracies like Russia and China (996). These phenomena, [Welzel \(2021\)](#) claims, are misinterpreted because scholars lack a “cultural theory of democracy” (1007). He defines democracy as “the idea that people live in freedom and have an equal voice and vote in the public sphere” ([Welzel 2021](#), 995) and makes the case for “the long-term reproductive supremacy of democracy over autocracy” ([Welzel 2021](#), 1011).

[Welzel \(2021\)](#) argues that emancipative values—first formulated by liberal philosophers—have spread globally since the Industrial Revolution, driven by improved living conditions and what he calls “cognitive mobilization,” i.e., the penetration of education, information, science, and technology into people’s lives (999). This process leads to a “psychological awakening” and a “cultural metamorphosis” as “collective mentality patterns change,” cultivating a culture in which individuals use “the faculty to act with purpose and to think for themselves” ([Welzel 2021](#), 994).

For [Welzel \(2021\)](#), these emancipative values do not require top-down implementation: “no civic education program and no orchestrated political strategy” (994) is necessary, because humans “by their very nature” cherish these values when living conditions enable them (999). Accordingly, democracy and democratic institutions are co-evolving with this cultural transformation: Democratic regimes institutionalize emancipative values, while those values legitimize and sustain democratic governance.

In Welzel’s view, a democracy sustained only by elites is unsustainable. Democratic stability occurs when institutions align with the cultural “demand” for freedom, while “democratization” results from a mismatch in which institutional supply lags behind cultural demand ([Welzel 2021](#), 1007). Cultural change drives political change, not the other way around ([Welzel 2021](#), 1009).

However, not all cultural forces support democracy. Nationalism and religion, [Welzel \(2021\)](#) argues, “certainly slow down” democratization (1002). Both are seen as preserving values “irreconcilable” with emancipative ones—namely hierarchy, obedience, and tradition—which serve to reinforce autocratic structures. In contrast, science and technology foster autonomy, critical thinking, and freedom. In short, in Welzel’s framework, science promotes democracy, while religion and nationalism threaten it.

21.3 Democracy is sustained by value judgments

[Welzel \(2021\)](#) contends that a democratic culture evolves “irrespective of a culture’s particular traditions” (1002). Democracy is essentially sustained by the enculturation of emancipative values. These values displace nationalism and religion since they embody the opposite of emancipative values. At the level of institutionalized democratic culture, nationalism and religion can have negative effects for democracy in accordance with the dynamics that drives the co-evolution of democratic culture and its institutionalization. Nationalism and religion serve only autocratic political regimes.

In [Jasanoff’s \(2005\)](#) discussion, we find compelling evidence that such a claim is mistaken, insofar as democratic decision-making requires not just the presence of values (such as emancipative ones), but value judgments—and it is unrealistic to assume that these judgments can be made without

drawing on substantive ethical, political, or religious traditions that extend beyond science alone. [Jasanoff \(2005\)](#), in her comparison of how Britain, Germany, and the United States responded to the challenges posed by the same biotechnologies, finds that “each democratic society tested, and to some extent reconstituted, its core understanding of what is political about science and technology, and how to organize and conduct the associated processes of technopolitics” (280). This observation highlights a commonality among all three countries.

However, “despite their structural similarities,” [Jasanoff \(2005\)](#) explains, “the three countries differ importantly in their legal and political traditions, with substantial implications for the conduct of politics” (30). According to [Jasanoff \(2005\)](#), nothing is “intrinsically political or apolitical by [its] very nature” (205). This definition coheres with [Brown’s \(2009\)](#) understanding of “politics” as “wherever people seek to mediate relations of conflict and power” (12). [Jasanoff \(2005\)](#) uses the term “civic epistemology” to refer to the publicly shared, culturally specific ways in which societies evaluate, validate, and use knowledge—especially scientific and technical knowledge—in governance and decision-making (247–71). It describes how different political cultures establish credibility, authority, and legitimacy in scientific knowledge, shaping the interaction between science, technology, and democracy. If Jasanoff is correct, then this constellation of differences can help illuminate the dynamics of value conflicts and the judgment they solicit in the evolution of democratic society.

At least three planes of development must be considered, according to [Jasanoff \(2005\)](#). They are as follows:

- a *The Political Shape of a Democracy*: While the United States, Germany, and Britain share structural similarities, such as the division of powers, each country’s democracy is shaped by a distinct political culture (see [Jasanoff 2005](#), 8, 21–23). Britain’s political culture, for example, is characterized by “the urge to create authorized bodies who can be trusted to see correctly for the people” ([Jasanoff 2005](#), 186) and the “national virtue of ‘muddling through’” ([Jasanoff 2005](#), 188). In contrast, the United States tends to resolve value conflicts “through litigation, in terms of individual rights” ([Jasanoff 2005](#), 164). Meanwhile, in Germany, [Jasanoff \(2005\)](#) highlights how “the supreme question of German identity” continues to be shaped by the trauma of “the twelve years of Nazi rule,” profoundly influencing its political culture (183). Stem cell research is not merely perceived as a promising avenue for developing effective medical treatments; rather, it is often viewed as a threat to vulnerable individuals, whose dignity may be compromised by state-sponsored medical research programs reminiscent of the Nazi-era “biocracy”—a system of governance in which biological and racial ideologies dictate political and social policies. The term is often used to describe the Nazi regime’s use of biological determinism and eugenics to justify the systematic persecution and extermination of certain groups, particularly Jews, disabled individuals, and other marginalized communities.
- b *Reenactment of Political Culture*: These differences largely explain the significant divergences in how each country has handled the same biotechnologies. The way biotechnologies are “framed” varies according to political culture (see [Jasanoff 2005](#), 42–67). In the United States, they are framed “as a collection of products” ([Jasanoff 2005](#), 95); in Britain, “as a potentially hazardous technological process”; and in Germany, “as a threatening program of state-sponsored control of society through technoscience.”

As a result, political cultures shape differing value judgments about biotechnology, which, in turn, define each country’s democratic response. Political culture encompasses various traditions that collectively shape value judgments, which, in turn, influence the very political culture from which they emerge. Moreover, value judgments are necessarily “multi-traditional” due to the nature of the questions they address. This is especially true in discussions about life. Only a

narrowly defined biological conception of life would justify restricting value judgments solely to the scientific tradition.

Controversies surrounding biotechnology serve as “sites of reenactment of political culture” (Jasanoff 2005, 143), reinforcing how societies interpret and respond to scientific and technological advancements. In this sense, “very different national strategies of normalization” emerged, Jasanoff (2005) argues, for making biotechnology “seem mundane and governable in the face of moral uncertainty and conflict” (147).

- c *Coproduction of Political Culture and Democratic Will Formation*: The value judgments about biotechnology do not merely reflect political culture; they also have the power to reshape it, influencing the evolution of democracy in each country. Jasanoff (2005) refers to this process as the “coproduction” of science and society, where scientific developments and social institutions shape each other. Numerous examples throughout Jasanoff’s (2005) work illustrate this dynamic.

The most salient example of *coproduction* at the European Union level occurred in the 1990s. “Europe’s distinctive framing of the risks of biotechnology in the 1990s,” Jasanoff (2005) argues, “emerges as one step in a larger project of creating order in science and society, or coproduction” (92). Several factors contributed to this process-oriented policy framing, including “the absence of well-developed regulatory regimes, a relatively passive industry,” the Green Party’s “influence in the Parliament,” the work of the Directorate-General for Environment and Nuclear Safety, and “a holistic discourse about the life sciences.” These elements combined to shape a uniquely European approach to biotechnology, one that, Jasanoff (2005) notes, “initially set European policy apart from directions taken in the United States” (92).

In turn, biotechnology played a key role in shaping European identity. Jasanoff argues that biotechnology helped create “issues, publics, and values that can be designated as truly European.” Much like nineteenth-century nation-states, “the EU has found it necessary to specify the problems it wants to solve in order to legitimate its political existence.” By the turn of the century, Europe had become “a space to watch for the deliberative politics of science and technology,” a development that, at least in part, resulted from the complex evolution of European biotechnology policy (Jasanoff 2005, 93).

In more established democracies, a compelling example of *coproduction* is the struggle over the ontology of the embryo in Britain following the birth of Louise Brown in 1978—“the first baby,” Jasanoff (2005) recalls, “produced through in vitro fertilization technology developed by British researchers Patrick Steptoe and Robert Edwards” (292). The ensuing debate over the moral status of the embryo in its first 14 days provides an illustrative case of “a neat display of coproduction,” Jasanoff (2005) argues, because the “adoption of new rules for ethical conduct of science resulted, at one and the same time, in the coming into being of a new natural entity” (152).

This entity, the “pre-embryo,” was defined as “a mass of undifferentiated cells” and was conceptually distinguished from the “embryo proper”—the authentic precursor of human life. This distinction was not merely “a claim of empirical observation,” Jasanoff (2005) reasons (153). Instead, it required “cultural authority, both religious and secular,” to withstand skepticism and to provide the pre-embryo with the ontological robustness needed to justify its use in scientific research. Ultimately, the “ontological and political reality” of the pre-embryo “was not a product of biological knowledge alone but was created (or coproduced) out of a complex mix of pragmatism, empiricism, and trust in experts” (Jasanoff 2005, 155).

In conclusion, insofar as it is true that democracy is rooted in emancipative values, from the perspective of Jasanoff (2005), their enculturation alone does not sustain democracy. Instead, it is value judgments that sustain democracy. Emancipative values surely feature in these judgments, but not exclusively—and probably never in a pristine form. While Jasanoff’s (2005) account is primarily

descriptive, it highlights that democratic decision-making is unavoidably shaped by a range of ethical, political, and religious traditions. In what follows, I turn to Habermas to argue that such shaping is not only inevitable but also normatively appropriate: A democracy that excludes these traditions from public reasoning risks impoverishing the very value judgments it requires. This includes the religious traditions that have shaped a democratic society. In short, a cultural theory of emancipative values does not necessarily translate into a secularization imperative—that one ought to eliminate religious traditions for the benefit of science and society on grounds of an inevitable conflict between faith and reason. Consequently, aligning science with democracy and religion with autocracy seems questionable.

21.4 The liberal state's democratic pluralism

Like Welzel (2021) and Jasanoff (2005), Habermas (2008) recommends, for analytical purposes, distinguishing between constitutionally fixed democratic institutions (ranging from parliament to bureaucracy) and an informal, more flexible domain that is equally significant for the reproduction of a democratic society. Welzel (2021) refers to this domain as the “cultural anchor” and a driver of “institutionalized democracy,” while Jasanoff (2005) labels it “civic epistemology” and sees it operating in a reciprocal relationship with democratic will formation. Whereas Welzel’s (2021) “cultural anchor” enhances the authority of science and advances secularization, Jasanoff’s (2005) “civic epistemology” is pluralistic, framing both science and religion as significant agents in the process of democratic will formation.

According to Habermas (2008), this is how it should be for the sake of the liberal state. In his work, the “public sphere” is equivalent to Welzel’s (2021) “cultural anchor” and Jasanoff’s (2005) “civic epistemology,” insofar as it identifies the realm of values and norms that are significant for the proper functioning of formal democratic institutions. According to Habermas (2019a, 2019b, 2023 [2019a]), a secularization imperative (that one ought to eliminate religious traditions for the benefit of science and society on grounds of an inevitable conflict between faith and reason) has valid points. First, it rightly assumes that systems of public knowledge and forms of social integration are intertwined and not independent matters. In other words, the secularization imperative is absolutely right that epistemological and social matters are entangled. Second, faith and reason are in conflict. However, Habermas (2019a, 2019b, 2023 [2019a]) argues that such a secularization imperative is mistaken in how it conceptualizes the relationship between religious traditions and modern science in the context of democratic will formation. Habermas rejects the view that modernity is the result of a displacement of religious traditions. Instead, he conceives of modernity as a process of increasing differentiation and largely facilitated by the adaptation of religious traditions.

His skepticism about a displacement-model of secularization aligns with Michael Polanyi’s position that “it is misleading to assert that science” is “based on a total rejection of” (Polanyi, Ziman, and Fuller 2000, 15) what Welzel (2021) identifies as the opposite of emancipative values—namely, values such as tradition and authority. “The authority of science is essentially traditional” (Polanyi, Ziman, and Fuller 2000, 15). Science, Polanyi clarifies, “can be pursued and transmitted to succeeding generations only within an elaborate system of traditional beliefs and values, just as traditional beliefs have proved indispensable throughout the life of society” (Polanyi, Ziman, and Fuller 2000, 18). Polanyi explicitly calls into question the possibility of absolute self-determination in terms of Welzelian emancipative values (Polanyi, Ziman, and Fuller 2000, 18).

A strict dichotomy of science vs. religion in terms of emancipative values and their opposites is further undermined by the fact that, as early as the inception of democracy as a desirable political ideal in the eighteenth century, religious traditions were resourceful enough to lend support to it (see Dunn 2005, 111–2). Steane (2018) concurs and argues that “to a very significant extent, committed

Christian discussion and effort has been the place which nurtured and gave to the world the notion of universal human rights, the practices of empirical science, secular democracy, universal education, and universal welfare” (251).

In short, [Welzel’s \(2021\)](#) concept of a “cultural anchor,” in which only emancipative values prevail, is a fabrication made possible by ignoring value conflicts—both among emancipative values themselves and between emancipative and autocratic values—as well as the value judgments required to resolve these conflicts. Additionally, it misleadingly aligns emancipative values with science (which is assumed to be defined solely by emancipative values and to perpetuate only those values) while associating their opposites with religion and nationalism (both of which are presumed to be defined exclusively by autocratic values and to sustain only autocratic values). The entanglement of science and religion, sanctioned by [Jasanoff’s \(2005\)](#) “civic epistemology” and [Habermas’s \(1991\)](#) “public sphere,” is meant to avoid false dichotomies for the sake of the liberal state and the future of democracy.

In this regard, [Norris and Inglehart \(2019\)](#) make a particularly relevant point. Drawing on empirical evidence, they argue that “social conservatism per se does not generate tensions” with emancipative values. “The crucial point is whether socially conservative values are combined with authoritarian predispositions” (450). In other words, there is no necessary isomorphic relationship between democracy and autocracy, on the one hand, and emancipative values and their opposite—conservative values—on the other.

“But,” one might ask at this point, using the words of [Feyerabend \(1982\)](#), “is there not a tremendous difference between science on the one side, and religion, magic, and myth on the other? Is this difference not so large?” (78). Feyerabend continues his list of rhetorical questions, “and so obvious that it is unnecessary to point it out and silly to deny it?” (78). According to [Feyerabend \(1982\)](#), these rhetorical questions contain three key assumptions:

Assumption A: scientific rationalism is preferable to alternative traditions. Assumption B: it cannot be improved by a comparison and/or combination with alternative traditions. Assumption C: it must be accepted, made a basis of society and education because of its advantages. (78)

“Assumptions A to C” are perfectly useful to define what scientism means and to make sense of the secularist imperative insofar as it is meant to follow from scientism. Scientism seeks to sever religious traditions entirely from the process of democratic will formation for the sake of proper reasoning and democracy. The argument is that religious traditions weaken the authority of scientific reasoning, which uniquely possesses the power to bring about consilience in the order of knowledge. This, in turn, disables scientific reasoning as a useful model for addressing the problems of social integration, including those posed by religious diversity. Social integration through a free exchange of reasons is jeopardized by religious diversity because, according to this view, religion is not fundamentally about reasoning but rather about irrational factors such as tribalism, emotions, and power. These factors are often cited to explain the occurrence of social conditions deemed undesirable from a democratic standpoint. Religion, in this sense, is seen as sand in the machinery of democracy.

Religion is thus portrayed as problematic in two ways. First, it poses a threat to reasoning—both in the claims it makes about the nature of things (such as explaining natural phenomena in supernatural terms) and in the manner it engages in reasoning within theological reflection (such as accepting beliefs on poor evidential grounds). Second, thereby it threatens social cohesion by fostering sectarianism, as its explanations conflict with scientific theories and thus disrupt consilience in the order of knowledge. In effect, threats to social cohesion materialize. These two threats—one to reasoning and one to social cohesion—are deeply intertwined, as democratic will formation relies on a willingness to give and accept reasons for the sake of overcoming differences and thereby to advance social integration.

Habermas (2008) articulates a strong disagreement with a secularization imperative (that one ought to eliminate religious traditions for the benefit of science and society on grounds of an inevitable conflict between faith and reason) along these lines, although he is a committed atheist and advocate of a strong version of the secularization thesis, which maintains that reason and faith are ultimately irreconcilable. To sketch his rather nuanced position, I will first present his definition of the liberal state and then introduce his reasons for the admissibility of religious arguments in the public sphere for purposes of democratic will formation in the liberal state.

Habermas (2008) defines the liberal state in terms of the following principles:

- 1 *The Principle of Enlightenment.* The state assumes that all citizens share the capacity to produce reasons without supernatural support and in the absence of esoteric assumptions. All citizens are equipped with “natural reason”—a defining notion of modernity that refers to the idea that human reason, unaided by divine revelation or religious authority, is capable of understanding the world, establishing moral and political norms, and guiding scientific inquiry.
- 2 *The Principle of Religious Liberty.* Individuals enjoy religious freedom as both a positive and a negative right. This means that individuals have the right to choose a religious identity (e.g., becoming or remaining a Christian) and, at the same time, the right to be protected from religious intrusion in their lives.
- 3 *The Principle of Legitimacy.* All exercise of state power must be justified to its citizens. This requires the presence of arguments that explain why power is exercised in a given way, rendering it, in principle, reasonable and acceptable to the citizenry.
- 4 *The Principle of Equal Participation.* Democratic will formation must be open to all citizens and must grant equal participation in deliberations and decision-making processes.
- 5 *The Principle of Worldview Neutrality.* The state must remain neutral with respect to competing worldviews. Habermas (2023 [2019a]) proposes “to apply the term ‘worldview’ quite generally to the picture that participants in a lifeworld form” (328). The lifeworld encompasses the knowledge, assumptions, and internalized beliefs that give meaning to individual actions and guide decision-making.

Habermas (2023 [2019a]) offers a phenomenological definition of the lifeworld through concrete examples:

The experience of fear of losing one’s footing in loose gravel or the sensation of blushing over an embarrassing mistake; how it feels when one suddenly realizes that one can no longer count on the loyalty of an old friend, or what it means for a long-cherished background assumption to be unexpectedly disappointed—we “know” all of these things. (326)

This embodied knowledge informs the picture of the world that participants in a lifeworld develop when forming a worldview. Consequently, the liberal state must refrain from adopting or endorsing any particular worldview, ensuring that it remains neutral in matters of belief and meaning as embodied in a lifeworld.

Habermas (2008) offers numerous reasons against eliminating religion from the public realm in democratic societies that operate under the *principles of enlightenment, religious liberty, legitimacy, equal participation, and worldview neutrality*.

The first argument is that in well-established constitutional states, churches and religious communities play a significant role in stabilizing and advancing liberal political culture. They contribute to political socialization, encouraging and motivating individuals to engage in political life. More specifically, religious communities support crucial forms of civic engagement, including volunteerism,

fundraising, fostering solidarity, and facilitating education. I refer to this as the *argument of political infrastructure*.

Second, it is in the liberal state's interest for religious citizens to participate in democratic will formation by virtue of their religious identity, as the state cannot afford to disregard potential sources of meaning and normativity. Religious traditions retain cognitive potential that liberal states have good reasons to value for the sake of their constitution and objectives. In particular, as [Habermas \(2008\)](#) observes, "religious traditions have a special power to articulate moral intuitions, especially with regard to the vulnerable forms of communal life" (131). I refer to this as *the argument of potentiality*.

Third, religious commitment permeates all aspects of a person's life. This reality makes it not only unrealistic but also impossible to expect religious citizens to compartmentalize their beliefs to the extent that they rely solely on secular reasoning in democratic will formation. This expectation is particularly unfeasible given that religion itself often has a deeply political dimension. In short, a secularization imperative for the sake of democracy within a liberal state would amount to demanding a form of self-denial from religious citizens. I refer to this as the *argument of integrity*.

Fourth, the liberal state can fully benefit from religious contributions to deliberation only if these contributions are translated in ways that make them accessible and acceptable to non-religious individuals, despite possible disagreements about the political issues under discussion. The task of translation should be a cooperative endeavor, in which both religious and non-religious citizens participate. Without such collaboration, religious citizens bear an asymmetrical burden in liberal democracies, as they are expected to handle the entire translation process while secular citizens simply await the results. This imbalance appears morally problematic. I refer to this as *the argument of cooperation*.

Fifth, religious citizens are fully capable of meeting the liberal state's expectations as outlined in the "institutional translation proviso," which requires that religious contributions in the public sphere be translated into non-religious arguments by the time they influence decision-making at the level of the political system—namely, in parliament, courts of law, and government. As [Habermas \(2008\)](#) argues, "Western culture has witnessed a modernization of religious consciousness as a response of religious traditions to the challenges posed by the fact of religious pluralism, the emergence of modern science, and the spread of positive law and secular society" (136). Theology plays a crucial role in this process, facilitating the evolution of religious traditions and theological insights for moral and political deliberation. In this sense, theology contributes to setting the pace for social progress. I refer to this as *the argument of ability*.

Sixth, non-religious citizens should adopt an epistemic attitude toward religion that goes beyond mere tolerance, as failing to do so reflects an inadequate self-critical stance toward enlightenment and modernity. What is expected of non-religious citizens, [Habermas \(2008\)](#) insists, is a mindset grounded in "a self-critical assessment of the limits of secular reason" (139–40). This approach serves as the non-religious counterpart to reflexive religious consciousness and embodies what [Habermas \(2008\)](#) describes as an "agnostic" yet "nonreductionist" form of "postmetaphysical thinking" (140). I refer to this as *the argument of critical enlightenment*.

While religious citizens, by virtue of their religious identity, have full access to democratic decision-making deliberations and enjoy the right to have their voices heard, the principles of the liberal state, [Habermas](#) insists, impose an absolute limit on the reach of religious contributions. "In a secular state," and "the liberal state is meant to be such a state," only "those political decisions can count as legitimate," [Habermas \(2008\)](#) asserts, "that can be impartially justified in light of generally accessible reasons" (122). This means that laws and policies affecting everyone must be justified equally for religious and non-religious citizens alike, as well as for citizens of different confessions. Any "exercise of power that cannot be justified in an impartial manner is illegitimate because it reflects the fact that one party is forcing its will on another" ([Habermas 2008](#), 122). More broadly, "citizens of a democratic polity are obliged to provide each other with reasons, because only in this way can

political power shed its repressive character” (Habermas 2008, 122). Thus, Habermas confines the admissibility of religious contributions to the “public sphere” only and not beyond to the level of the political regime (such as government), unless they are translated into secular reasons.

Despite this considerable restriction on the role of religious contributions in democratic will formation, their permitted inclusion remains significant, because Habermas (2022) maintains that “the function of the public sphere” is to ensure “the continued existence of the democratic political community” (2). The public sphere, Habermas (2022) explains, “within the functionally differentiated structure of modern societies,” occupies a space “between civil society and the political system” (2).

It is important to note that the public sphere is essential to a Habermasian deliberative democracy, which is defined by the principle that the exercise of state power in democratically constituted societies gains legitimacy not merely through the division of power, due process, voting, and proper representation, but through the quality and inclusiveness of the deliberative process. This process results in laws and policies that guide the exercise of state power. State power is legitimate if it is derived from the discourse’s ability to reflect the collective will and the shared understanding of the participants. The public sphere serves as the primary means for citizens to refine arguments emerging from civil society, thereby fostering an informed public that can hold political institutions accountable.

The notion of deliberation, which features prominently in Habermas’s theory of democracy, is a sociological category that is not “sharply defined,” as Habermas (2022, 62) explains. It aims to capture “from an observer’s perspective” what occurs when citizens engage in discourse for the purpose of democratic will formation.

From a participatory perspective, it becomes evident that political arguments are not merely about specific issues (such as abortion, gay marriage, climate action, militarization for peace through strength); they also embody epistemological standards and ideals. While political discourse is certainly not akin to “a convivial university seminar,” Habermas (2022) acknowledges, it remains true that “to argue is to contradict” (64). Contradiction through argument “elicits the epistemic potential of language without which we could not learn from one another” (Habermas 2022, 64). This, he explains, is the essence of deliberative politics: “by engaging in political disputes,” Habermas (2022) argues, “we improve our beliefs and thereby approach the correct solution to problems” (64). Through this process, cooperation among citizens is strengthened, and the social integration of the citizenry is advanced. Effective communication is crucial in this regard, as it serves the function of communicative action, which advances social integrity.

21.5 Conclusion

The political ideal of democracy that people live in freedom and have an equal voice and vote may be seen to be rooted in emancipative values. The enculturation of these values gives rise to value conflicts, however, that require judgments shaped by the particular traditions influencing the political culture of a democratic society. A liberal state, and I considered only the situation in liberal states following Habermas, has not only a genuine self-interest in treating scientific and religious perspectives equally in the public sphere as it seeks normative resources to strengthen its foundation and operations, but it can also expect that mutual learning between secular and religious citizens will enhance democracy. *The argument of critical enlightenment* is particularly poignant in this respect.

Since the public sphere is pivotal for the reproduction of the liberal state, the inclusion of religious traditions within it is significant. Religious traditions have developed theologies that function as catalysts for social progress, comparable to the philosophical work on scientific theories concerning the universe and humanity’s place within it. A scientific secularization imperative risks undermining democracy by treating religious arguments unequally. Moreover, such an imperative threatens the liberal state, as it risks violating the principle of worldview neutrality by favoring *scientific* worldviews.

References

- Bronowski, Jacob. 1972. *Science and Human Values*. Harper & Row.
- Brown, Mark B. 2009. *Science in Democracy: Expertise, Institutions, and Representation*. MIT Press.
- Dunn, John. 2005. *Democracy: A History*. Penguin Canada.
- Ecklund, Elaine Howard, and Elizabeth Long. 2011. "Scientists and Spirituality." *Sociology of Religion* 72 (3): 253–74. <https://doi.org/10.1093/socrel/srr003>.
- Feyerabend, Paul. 1982. *Science in a Free Society*. Verso.
- Habermas, Jürgen. 1991. *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*. Translated by Thomas Burger. MIT Press.
- Habermas, Jürgen. 2008. "Religion in the Public Sphere: Cognitive Presuppositions for the 'Public Use of Reason' by Religious and Secular Citizens." In *Between Naturalism and Religion: Philosophical Essays*. Translated by Ciaran Cronin. Polity.
- Habermas, Jürgen. 2019a. *Auch eine Geschichte der Philosophie: Band 1: Die okzidentale Konstellation von Glauben und Wissen*. Suhrkamp.
- Habermas, Jürgen. 2019b. *Auch eine Geschichte der Philosophie: Band 2: Vernünftige Freiheit: Spuren des Diskurses über Glauben und Wissen*. Suhrkamp.
- Habermas, Jürgen. 2022 *A New Structural Transformation of the Public Sphere and Deliberative Politics*. Translated by Ciaran Cronin. Polity.
- Habermas, Jürgen. 2023 [2019a]. *Also a History of Philosophy: Volume 1: The Project of a Genealogy of Post-metaphysical Thinking*. Translated by Ciaran Cronin. Polity.
- Jasanoff, Shelia. 2005. *Designs on Nature: Science and Democracy in Europe and the United States*. Princeton University Press.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Lafont, Cristina. 2022. *Democracy without Shortcuts: A Participatory Conception of Deliberative Democracy*. Oxford University Press.
- Norris, Pippa, and Ronald Inglehart. 2019. *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*. Cambridge University Press.
- Polanyi, Michael, John Ziman, and Steve Fuller. 2000. "The Republic of Science: Its Political and Economic Theory Minerva, I (1) (1962), 54–73." *Minerva* 38 (1): 1–32. <http://www.jstor.org/stable/41821153>.
- Russell, Bertrand. 1932. *Education and the Social Order*. Unwin Hyman Limited.
- Russell, Bertrand. 1939. "Education and Democracy." *Bulletin of the Department of Secondary-School Principals of the National Education Association* 23 (81): 6–16. <https://doi.org/10.1177/019263653902308101>.
- Steane, Andrew M. 2018. *Science and Humanity: A Humane Philosophy of Science and Religion*. Oxford University Press.
- Welzel, Christian. 2021. "Democratic Horizons: What Value Change Reveals about the Future of Democracy." *Democratization* 28 (5): 992–1016. <https://doi.org/10.1080/13510347.2021.1883001>.

VALUES IN SCIENCE IN NONDEMOCRATIC CONTEXTS

Karoliina Pulkkinen

22.1 Introduction

Much has been said on science, values, and democracy (e.g., [Douglas 2009](#); [Kitcher 2001, 2011](#); [Le Bihan 2024](#); [Lusk 2020, 2021](#); [Schroeder 2017, 2021, 2022](#); [Thoma 2024](#)). Science in nondemocratic settings, however, has not received as much attention beyond a few exceptions in the philosophical literature on science ([Fritsch 2023](#), 1011–4; [John 2019](#)). The aim of this brief chapter is to motivate a closer look into values in science beyond democratic orders. Aside from pointing toward new or unexplored topics, we can gain new ways to lean into existing themes such as ideological influences of values in science. A closer look into such settings can also help draw attention to some underlying assumptions in the current debates. However, a closer look into nondemocratic orders will show that there are also many aspects that are similar to democratic settings. For this reason, a call to broaden the scope should be accompanied with a readiness to question such quests insofar as they are motivated by the need to paint them as inherently different from democratic orders. That remains something to be shown rather than something to be assumed from the get-go.

First, some disclaimers. Labels like “authoritarian contexts” or “authoritarian countries” carry a risk of being reductive in the same way as exploring, say, the 1970s US climate science solely in relation to the political order of the day would be reductive. Furthermore, such labels hide a great variation, plausibly encompassing eighteenth-century enlightened absolutism and contemporary countries ruled by authoritarian regimes. My discussion is constricted in the sense that it draws from the works of historians of Soviet science and technology. Although at a first glimpse, it might look like this poses a limitation, there are extremely good reasons to adopt the Soviet science system as a starting place. Aside from providing a pool of cases on how something that might be labeled “Soviet” or “Marxist” values influenced research and its organization, a major reason to investigate it concerns the importance attached to science. During its existence, the Soviet Union established a formidable science system which organized much of the research into a substantial network of research institutes that were geared toward freeing scientists and engineers from teaching. As noted by historian of science Alexei [Kojevnikov \(2004\)](#), its achievements put some pressure on the Popperian, post-World War II wisdom that democracy is necessary for thriving of science and technology. Furthermore, literature on Soviet science—and histories of the Soviet Union more broadly—include many methodological reflections that may offer some guidance in broadening the scope of studies to cover nondemocratic orders.

To briefly advertise, I identify three insights that can be gleaned from looking more closely at values in science in nondemocratic settings. First, a closer look into historiographical debates challenges the assumption that power in authoritarian settings flows seamlessly in a “top down” manner (Section 22.2). Second, historical accounts can broaden our understanding of ideological influence of values on science, where greater attention is needed for understanding ideological impact on public discourse, as philosopher Stephen John (2019) has already intimated in his take on Lysenkoism as wishful speaking (Section 22.3). Finally, the role of science and values in maintaining nondemocratic political legitimacy—particularly political legitimacy gained through the successful obtaining of important-deemed outputs—is something that should be incorporated in philosophical analyses concerning science, values, and political legitimacy (Section 22.4). Section 22.5 concludes.

22.2 From a simple binary to a complex Möbius strip

One of the assumptions regarding authoritarian orders is that political power is wielded from top down where other social groups may respond by complying or resisting such endeavors. This assumption is reflected in discussions entertaining the idea that authoritarian orders might be more efficient at tackling some persistent global problems than more pluralist democratic societies (see e.g., the discussion in Fritsch 2023, 1012–3). Correspondingly, when looking at Soviet science, the expectation may be that Soviet politicians imposed their preferred values on research: If researchers then failed to comply, they were penalized accordingly. This line of reasoning is, for example, present in some Western commentaries regarding the role of dialectical materialism in scientific research (as noted in Graham 1987).

However, one of the most interesting aspects of Soviet histories concerns how difficult and fraught it is to determine just how top-down political power was. A good counterexample concerns the very features of the Soviet science system which were put to place shortly after the revolution of 1917, largely in collaboration with Bolshevik revolutionaries and scientists and other technical specialists (Bailes 1978). Some of its basic features include the fact that research was state sponsored. Politicians expressed a great appetite for applied research that was geared toward state-building and embraced the idea of organizing research around research institutes (Josephson 1988; Kojevnikov 2004, chap. 1). Although all these aspects owe to political initiatives, a great many researchers—many of whom held very different political views to those of Bolsheviks—nevertheless welcomed the new resources and opportunities. One element of consensus included the idea of organizing research around independent research institutes instead of universities, which found its precedent in comments voiced by scientists prior to the revolution (Adams 1980; Brooks 1997, 176). Similarly, the idea of bolstering applied research found its support after the World War I, as the awareness of the lack of the Russian Empire’s development on this area grew more acute (Kojevnikov 2002). Thus, many of the systemic aspects of Soviet science were initially put in place in collaboration with politicians and scientists, with Bolsheviks often latching onto ideas that had been circulating prior to 1917.

To secure access to expertise, Bolshevik revolutionaries did not initially exert much control over the kinds of research conducted, although they incentivized research of more applied nature. It is especially in the subsequent period of Stalin’s rule that the top-down conceptualizations of power initially appear more supported. Stalin’s “Great Break” of 1927, which encompassed profound social, economic, and political transformations, paved the way to much tighter administrative control of science and academia, together with more centralized science planning (Kolchinsky 2014, 332; Roll-Hansen 2004, 92–4).

The apparent appropriateness of the more top-down conceptualizations of political power is at times demonstrated with the rise of Trofim Lysenko’s (1898–1976) “agrobiology” or “Michurinism,”

which came together with the decline of genetics. Lysenko, a member of the new generation of technical specialists who had benefitted from the educational policy set by the Bolsheviks, was initially supported by the established plant scientist Nikolai Vavilov (1887–1943) who believed that state-funded science could help to mobilize agricultural sciences to overcome famines. Initially Vavilov, who was the leader of the Institute of Plant Breeding (VIR), the president of the All-Union Academy of Agricultural Sciences (VASKhNIL), and the director of the Institute of Genetics, had a good working relationship with Bolshevik science administrators. Especially after Stalin’s “Great Break,” however, he suffered from accusations of failing to gear science for the socialist reconstruction of agriculture (Kolchinsky 2014, 333). In contrast, Lysenko, working together with the philosopher Isaac Prezent, formulated a “Michurinist” version of his agrobiology that was especially well-aligned with practical goals and appropriately highlighted the domestic achievements of Russian practitioners. Lysenko managed to gain the leadership of virtually all key institutions previously held by Vavilov, whereas Vavilov, who was imprisoned in 1940, died in a Saratov prison in 1943—a fate that he shared with a number of his closest associates (Kolchinsky 2014, 350; see also Gordin 2022).

The first accounts of Lysenkoism that emerged after the widespread anti-Lysenko campaigns of the late 1940s and 1950s tended to depict Lysenko’s rise as a direct result of the coercive totalitarian system that controlled every facet of Soviet life. In these accounts, the chief reason for the demise of genetics was its incompatibility with the doctrines of Marx and Engels, which was something that Lysenko’s views succeeded at establishing, thus giving an example of communist, Marxist biology (Graham 1987; Krementsov and deJong-Lambert 2017, 13–4). The historical accounts that depicted Lysenkoism in such terms are representatives of a broader historiographic tendency known as the Cold War-inspired “totalitarian model.” Often associated with Hannah Arendt’s work on totalitarianism, this model tended to side-step the scientific achievements of the Stalinist era and depicted Soviet science “as a victim of pervasive control by the Party/state apparatus” (Gerovitch 2002, 5). Originating from the field of Soviet studies that were largely dominated by political scientists, the totalitarian model depicted the Soviet Union in top-down terms, characterized by a near-total control by the state. A central theme was the conflict between state and society, which left autonomous associations destroyed, bonds between people atomized, resulting in a “a powerless, passive society that was purely an object of regime control and manipulation” (Getty 2008, 713; direct quote from Fitzpatrick 2007, 80; see also Fitzpatrick 1986).

In the 1970s, as more social historians entered the field previously dominated by political scientists, the initial Cold War premises of the debate were revised. The “new cohort” criticized the totalitarian model for being politically biased and inappropriate at depicting contemporary Soviet reality. They found it particularly difficult to endorse the view of society without internal tensions and conflicts (Fitzpatrick 1986, 357, 361; see also Kotkin 2001). Because of this, the new “revisionist” approach explicitly challenged the prevailing totalitarian assumptions of society as “irrelevant to an understanding of Stalinist political processes” and depicted the power relations in more of a bottom-up way (Fitzpatrick 1986, 370; 2007, 81). Reflecting this, for example, historian Kendall Bailes argued that adopting “group conflict models” alone failed to capture the complex relationships between scientists, engineers, and politicians. Alongside it, Bailes advocated using a model that focuses on the “interaction of social forces,” including forces of social cohesion (1978, 8).

By the formative 1990s, there was a notable retreat from the use of the two labels “totalitarian” and “revisionist,” reflecting the search for alternative frameworks (Kotkin 2001, 709). In endorsing a Foucauldian perspective, the “post-revisionists” challenged both the “from below” and “from above” approaches as equally problematic and preferred construing power relations as deciphered and multiple (Fitzpatrick 2007, 87). Where revisionists had been averse to questions of ideology, post-revisionists returned to the theme, although their understanding of it was “not in the sense of

a canonical body of texts ... but more as *Weltanschauung*—something collectively construed rather than imposed” (Fitzpatrick 2007, 87; see also Kotkin 1998). This was reflected in histories of science as historians started to “emphasize various forms of ideological accommodation, pragmatic cooperation, and even institutional integration between different groups of scientists and politicians” (Gerovitch 2002, 5). Furthermore, Western depictions of Soviet science as something different, or exceptional, began to be questioned, leading to examinations of Soviet science as “any set of events in the history of science in any country” (Adams 2001, 260).

Relating the above insights into the discussions on values in science gives the following. As noted earlier, the usual assumption regarding authoritarian orders is that power flows unobstructed from top down. This raises expectations of their efficiency in achieving goals, particularly tackling issues related to pressing global problems such as climate change. For example, Mittiga (2022, 999) argues that given the democratic publics’ resistance to climate action, “liberal-democratic governments (and theorists) must confront the bleak possibility that responding to the existential threat of climate change at this late stage may require ... embracing authoritarian power” (see also Fritsch 2023). My above discussion highlights that assumptions of the greater effectiveness of autocracies may employ a simplified conception of broader society as free of internal tensions. Insofar as one considers the efficient tackling of global challenges more pressing than maintaining pluralist democratic societies, embracing Mittiga’s bleak conclusion, in my view, requires further investigation and evidence, including rigorous comparisons with how pluralist democracies fare with similar challenges. In short, such assessments represent a starting point for investigations rather than their conclusion. A related observation is that scientists working in nondemocratic contexts can at times have a greater level of agency than is standardly thought due to the relatively powerful position of those with technical training and the potentially symbiotic relationship between science, technology, and political actors who facilitate or sponsor research.

Although with explorations of Soviet pasts, there is evidence to back up both conceptualizations of power in “top down” and “bottom up” terms, it is also worth noting that as frameworks, they are unspecific and broad and may lend themselves better for certain kinds of phenomena than others. With that caveat in mind, it appears that compelling analyses of nondemocratic contexts call for an ability to focus on these opposing aspects at once—a kind of a chameleonic vision instead of that of a cyclops.

This observation regarding our underlying assumptions on political power have some implications for philosophical analyses that touch upon scientists’ agency to include the values they see important (both in democratic and nondemocratic settings). A good case to consider in relation to this is ideological influence on science. The more recent historical scholarship suggests that ideological influence on science was not necessarily imposed from above, but something that could be co-created with scientists, which raises questions on how to best understand such influence.

22.3 Ideology

Histories of the Soviet Union offer multiple ways to make sense of ideological influence on science. Perhaps the most misunderstood approach, however, is to treat Marxism and dialectical materialism as a rigid, codified system of belief that distorts content of science to align it with its core principles. This somewhat pejorative understanding of ideological influence does not account for instances when Marxism has played a positive, helpful, or harmless influence on research (Gramham 1987; Kojevnikov 2004, chap. 3). Furthermore, when applied to the rise of Lysenkoism, for example, it often goes unnoted that dialectical rhetoric was employed by many geneticists who eventually lost their institutional foothold to Lysenko (Adams 1980, 174; Roll-Hansen 2004, 178).

This gives rise to the question of why their accounts feature much more rarely as examples of ideological influence of values on science.

There are many other routes toward understanding ideological influence beyond the one represented above, though. Historian Mark B. Adams (1980, 173), who defined ideology as “bodies of publicly and officially articulated ideas or doctrines relating to policy, ideas, and action,” singled out four modes of ideological influence on science. First, ideology can play a *legitimizing* role by allowing new justifications for scientific research or ideas (without necessarily having direct effect on the research process itself). It may then also act as a *selector* in favoring or disfavoring certain lines of research, technique, theories, and fields of study. It can also play the role of *shaper* by inspiring new lines of research, experimentation, and theories that emerge or relate to ideological ideas (Adams 1980, 174–5). What will be further explored here is the notion of linguistic or discursive understanding of ideology (which, in my view, can be viewed to underpin all the above senses of ideological influence). Understood in this way, ideological influence can bridge the gulf between scientists’ technical work and the broader public priorities and values, including those who act as the patrons of science (Adams 1980, 174). Such discursive influence can be thought as giving a prominent or preferred frame or narrative, or a set of frames and narratives (see, e.g., McKaughan and Elliott 2013).

This discursive approach is given an especially developed treatment in Gerovitch’s (2002) account of cybernetics in the USSR, which examines its development as series of encounters between the Soviet ideological language and the language of cybernetics. Taking seriously the idea of ideological influence as a form of discursive practice, Gerovitch reutilizes the category of “newspeak” that featured originally in George Orwell’s 1949 novel *Nineteen Eighty-Four* to depict how Soviet language incorporated a totalitarian world view and the “mental habits” of Party members (Gerovitch 2002, 12–3).

Instead of viewing newspeak as reality-distorting ideologization of language, or solely a pragmatic instrument of negotiation, Gerovitch promotes a third alternative: Viewing it as a pervasive “cultural medium” that conveys scientists’ gradual adaptation “to the behavioural patterns, modes of thought, and rhetorical styles of their patrons.” Following historian Stephen Kotkin’s (1995) depiction of Stalinism as a “socialist civilisation,” which was accompanied with its own way of speaking (“Bolshevik”) that provided the speaker both access to power, a sense of identity, and a set of values, Gerovitch (2002) argues that

Academic discussions were, more and more often, conducted in a language accessible to the bureaucracy: the value-laden ideological language of official Soviet discourse. This language constituted a pervasive ideological-linguistic medium in which Soviet scientists lived and worked. (12)

The closeness between the language of science and politics suggests that newspeak was not a linguistic practice imposed on scientists. Rather, it represented a set of discursive strategies that could be used to adapt to the prevailing political situation and manipulated to scientists’ advantage (2002, 9).

This interpretation highlights that newspeak could be a malleable medium. It lent itself to negotiations between the scientists, the Party, and the government. Instead of simply expressing a set of orthodoxy, its skillful use enabled scientists to “manipulate ideology and, to some extent, to define what was permitted in a Soviet context.” In this sense, newspeak was simultaneously “manifestly dogmatic” and “strikingly flexible,” and hence it “did not embody the truth; it provided mechanisms for negotiating the truth.” In other words, it was not solely repressive but also productive as it could give scientists the tools to present their ideas and discredit those of their opponents (2002, 14).¹

Among philosophical literature on science, there have already been numerous advocations for a more central role to scientists’ communication (e.g., Elliott 2017, 2020; Franco 2017, 2019; McKaughan and Elliott 2013). To use the language endorsed by philosophers Daniel McKaughan and

Kevin Elliott (2013; see also Elliott 2020), one way to think of the discursive influence of ideology is to take it as a preferred and prominent—and at times, fuzzy—“frame.” The challenge for scientists, then, is to find ways to use this broader frame so as to allow maintaining a functional relationship with the (sole) patrons of one’s work while also holding onto one’s epistemic and ethical commitments. The example of the development of many fields of humanistic inquiry in the USSR helps to illustrate the difficulty. Due to the political pressures, many humanities scholars restored into more “factological” or “textological” lines of research, which emphasized the presentation of primary source materials without much accompanying interpretation or analysis. Steering away from interpretations that could be used as a political weapon—particularly after the difficult years of late 1940s—was for many historians a solution to continue research without compromising one’s professional integrity (Aronova 2011, 94, fn.58; Gerovitch 1996, 106, 1998, 199). This implies that difficulties and risks in framing one’s research could lead to the eschewing of any kind of frame.

From the existing philosophical accounts, the closest treatment to the above “newspeak”-style ideological influence comes from John (2019). Considering especially the strong form of the doctrine of *partiinost*, where scientists’ “epistemic commitments are guided by the Party’s assessments of what commitments they should hold,” John argues that its influence on science is better understood not in terms of wishful thinking but as wishful *speaking* (2019, 65). In part, this proposal is guided by the observation that scientists make scientific claims without necessarily believing them (Elliott and McKaughan 2014, 6; see also Elliott and Willmes 2013; McKaughan 2007). Since the idea of wishful thinking does not apply conveniently to scientists’ assertions that come without belief, it is useful to think of them in communicative terms. In contrast to wishful thinking, wishful speaking involves making claims that are not well-established where this is motivated by the non-epistemic benefits that follow from others believing (or believing that they believe) that claim, regardless of the claim’s truth (John 2019, 65–6). Irrespective of the quality of its consequences, John argues that all cases of it remain ethically problematic on the grounds that they disrespect their hearer. This is because

in speaking wishfully, a speaker treats hearers’ beliefs as mere means to be manipulated for the sake of the ends which the speaker values. Even if these ends are noble, such an attitude is to disrespect hearers’ status as autonomous agents.

(2019, 67)

Although the Soviet context illustrates especially well instances of wishful speaking, it also raises some complications for the above ethical assessment. For one, sometimes the alternatives to wishful speaking are not any less ethically problematic. This is because refraining from wishful speaking can be costly, too. For example, one might defend Einstein’s physics against political accusations of its idealism by claiming that such scholarship is Soviet in virtue of it being the same age as the revolution (see, e.g., Hall 2008, 233). Although asserting a genetic fallacy (without really endorsing it) may indeed be disrespectful toward the hearer, similarly refraining from defending physics from such accusations could be risky and contribute to its downfall. This implies that wishful speaking is in some circumstances a defensible maneuver. Furthermore, when both the hearer and the speaker know that the speaker is making unfounded claims for non-epistemic benefits, it is not clear that the issues of disrespect arise in the same way as when the hearer is unaware of the game.

All this is to say that articulating scientists’ normative responsibilities in settings that are saturated with wishful speaking is extremely challenging. Nevertheless, the above vignette on newspeak suggests that a good grasp of the prevailing ideology can provide scientists the discursive tools that can be used strategically in various ways to defend what they deem worth defending. Moreover, wishful

speaking can be morally permissible if the costs of refraining from it are significant enough. The thorniness of this issue is well illustrated by [Kojevnikov's \(2004, 159\)](#) poignant question in his account of the life of Nikolai Vavilov's brother Sergei Vavilov (1891–1951), a high-ranking official and one-time president of the Soviet Academy of Sciences, who indeed defended the “new physics” of Einstein and Bohr in the above ways. Kojevnikov notes,

The complexity of moral challenges of life under a dictatorial regime offered a sharper form of the universal and everlasting question: What does it mean to be moral in the immoral world? The choice then was often between open protest, which would also harm innocents, and compromise with what one considered as a lesser evil.

It would be beneficial to give greater attention to different forms of ideological influence—whether understood in terms of newspeak, legitimation, or as shaping and selecting of appropriate lines of research—and particularly how they can influence different disciplines, research, and the broader public. At best, a firmer empirical grasp can provide some further guidance on how to develop normative philosophical accounts concerning values.

22.4 Political legitimacy

In the contemporary discussions on policy-relevant research, the influence of scientists' values on research is at times viewed as a problem to remedy. As research could impact the broader public, it is often perceived as an issue if scientists, who are not democratically elected, have their values shape policy-relevant research. One way to frame this is to think of the problem in terms of democratic legitimacy of values (e.g., [Intemann 2015](#); [Lusk 2020, 2021](#); [Schroeder 2017, 2021, 2022](#); [for a critique Hilligardt 2023](#)). To alleviate the apparent lack of democratic legitimacy, the usual recommendation is to incorporate the values of the public into the research process, although there are some risks in doing so ([de Melo-Martin 2024](#)). It has however been noted that democratic legitimization of individual value-judgments in science may not necessarily increase overall legitimacy. This is because conflicts between different public institutions can be beneficial for the overall legitimacy of the democratic system ([Hilligardt 2024, 45](#)).

The increasing emphasis on political legitimacy raises questions about science conducted in non-democratic contexts such as the Soviet Union. In some sense, it is easy to recognize that the democratic solutions to legitimize values can be implemented irrespective of the overall political system insofar as scientists endorse the reasoning (and the ideals) animating the abovementioned democratic proposals. Despite this, there still remains the question of whether alternative notions of political legitimacy may be active in nondemocratic contexts.

Here, the Soviet experience is especially informative. When in the current contemporary proposals democratic theory serves as a way to legitimize aspects of scientific research, a closer awareness of Soviet history indicates a different order of direction: Namely, that science and technology were used to legitimize the regime and its political actors. A classic example of this is Stalin's use of the Soviet world records in aviation to bolster his legitimacy, which has prompted historian Kendall [Bailes \(1976, 1978, chap. 14\)](#) to argue that technology could play a legitimizing role for a developing society such as the Soviet Union. Values guiding research and research institutions form a part of this dynamic as is vividly illustrated by [Bailes' \(1976, 67–9\)](#) discussion of aviation designers' prioritization of specific features in the design of aircraft; emphasis on specific values could support or hinder broader legitimization ([Pulkkinen 2025](#)). Also, the Soviet space program served the purpose of legitimating the Soviet political order which highlights that legitimacy had an international dimension ([Gerovitch 2011](#)).

Due to its reliance on achievement of results that have been deemed important, this mode of legitimization can be analysed in terms of “output-legitimacy,” although other labels have been used, too (Gill 2011). To enhance our understanding of how science and technology can play such a legitimization function, there are at least three available approaches to pursue. First is to explore the normative appeal of output legitimacy in light of experiences both in the Soviet Union and other contexts. Some have already, for example, argued that output legitimacy is a part of normative democratic legitimacy when considering legitimacy of international organizations (Steffek 2015). Others have argued that more Weberian “descriptive” notions of political legitimacy, which are standardly excluded from normative discussions, do however retain “moral significance” under some circumstances (Greene 2017).

Aside from such more normative explorations, the remaining strategies include more naturalistic approaches that have an affinity with investigative modes of History and Philosophy of Science (HPS) and historical epistemology. For example, one could historicize the very notion of output legitimacy by mapping out how science and technology have been used to legitimize political projects and political actors. Aside from providing a better explanation or a description of such processes, also the notion of normativity of legitimacy could be made into an object of inquiry by studying how and whether various actors found different legitimization strategies convincing. These options gain some support from the fact that Bernard Williams (2005) considered real-life attempts to gain political legitimacy as an important aspect of broader philosophical understanding of political legitimacy. Furthermore, some contemporary theorists of political legitimacy—most notably Fabienne Peter (2023)—have argued that expertise (or in Peter’s terminology, “cognitive epistemic authority”) can be a source of political legitimacy, where such proposals would benefit from concrete examples arising from Soviet history to better understand their appeal and limitations. Irrespective of which approach is adopted, turning attention to the dynamics of political legitimization in the Soviet Union allows for a helpful way to broaden our understanding of political legitimacy and its relationship with science and technology.

22.5 Concluding remarks

As the above discussion indicates, there are some additional insights to be gained by turning attention to science in nondemocratic countries. We saw that historiographical developments problematize the idea that there is a seamless, unobstructed flow of top-down authoritarian power. This discussion also highlighted that scientists and other technical experts, as a group, were in a relatively powerful position due to their expertise and special skills, to the point that treating politicians and scientists as entirely different groups can be challenging. In the words of Gerovitch, the closer we look, the more complex the relationship between Soviet scientists and Soviet politicians appears: “Instead of a simple binary opposition, we have a confusing Möbius strip: it is no longer entirely clear who is on which side” (Gerovitch 2002, 4–6; see also Gerovitch 1996).

Aside from this, a closer consideration of Soviet science pointed toward new ways of understanding ideological influence and broadening our understanding of political legitimacy. Of course, my discussion just scratches the surface, but hopefully it is indicative enough of the potential that there remains in broadening the scope of our discussions. No doubt many other promising avenues of new research are there to be explored.

Acknowledgments

I remain grateful for the discussions at Perspectives of Science seminar at Helsinki (May 2024), BSPS (York 2024), SEPOS Work in Progress seminar at Michigan State University (September 2024), and SFY Helsinki (March 2025). Thanks are also due to ARGIH group at Aleksanteri Institute, Julia LaJus, Arnon Levy’s group, and the IUHPST Joint Commission for establishing the essay competition

that prompted some of the material of this essay. The feedback of the editors of this volume was invaluable in developing my position. The writing of this chapter was supported by a Kone Foundation grant for the project “Tracking the role of values in Soviet science.”

Note

- 1 Despite the level of agency permitted by newspeak, it is important to note that mastering it did not allow for a straightforward terrain of operation as political priorities could sometimes contradict. This is demonstrated by Stalin’s conflicting demands that Soviet science should not only to overtake but to surpass the achievements of science of other countries, which ran against the other requirement of the Party ideologies to treat Western scholarship as “idealistic and reactionary” in the “home front” of the Cold War. See discussion in Gerovitch (2002, 15–6).

References

- Adams, Mark B. 1980. “Science, Ideology, and Structure: The Kol’tsov Institute, 1900–1970.” In *The Social Context of Soviet Science*, edited by Linda L. Lubarno and Susan Gross Solomon. Westview Press.
- Adams, Mark B. 2001. “Networks in Action: The Khrushchev Era, The Cold War, and The Transformation of Soviet Science.” In *Science, History and Social Activism: A Tribute to Everett Mendelsohn*, edited by Garland E. Allen and Roy M. MacLeod. Kluwer.
- Aronova, Elena. 2011. “The Politics and Contexts of Soviet Science Studies (Naukovedenie): Soviet Philosophy of Science at the Crossroads.” *Studies in East European Thought* 63 (3): 175–202. <https://doi.org/10.1007/s11212-011-9146-y>.
- Bailes, Kendall E. 1976. “Technology and Legitimacy: Soviet Aviation and Stalinism in the 1930s.” *Technology and Culture* 17 (1): 55–81. <https://doi.org/10.2307/3103253>.
- Bailes, Kendall E. 1978. *Technology and Society under Lenin and Stalin: Origins of the Soviet Technical Intelligentsia, 1917–1941*. Princeton University Press.
- Brooks, Nathan. 1997. “Chemistry in War, Revolution, and Upheaval: Russia and the Soviet Union, 1900–1929.” *Centaurus* 39: 349–67.
- de Melo-Martin, Inmaculada. 2024. “Concerns about Contextual Values in Science and the Legitimate/Illegitimate Distinction.” *Philosophy of Science* 91 (4): 851–68. <https://doi.org/10.1017/psa.2024.20>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2020. “Framing Conservation: ‘Biodiversity’ and the Values Embedded in Scientific Language.” *Environmental Conservation* 47 (4): 260–8. <https://doi.org/10.1017/S0376892920000302>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2014. “Nonepistemic Values and the Multiple Goals of Science.” *Philosophy of Science* 81 (1): 1–21. <https://doi.org/10.1086/674345>.
- Elliott, Kevin C., and David Willmes. 2013. “Cognitive Attitudes and Values in Science.” *Philosophy of Science* 80 (5): 807–17. <https://doi.org/10.1086/673719>.
- Fitzpatrick, Sheila. 1986. “Afterword: Revisionism Revisited.” *The Russian Review* 45 (4): 409–13. <https://doi.org/10.2307/130471>.
- Fitzpatrick, Sheila. 2007. “Revisionism in Soviet History.” *History and Theory* 46 (4): 77–91. <https://doi.org/10.1111/j.1468-2303.2007.00429.x>.
- Franco, Paul L. 2017. “Assertion, Nonepistemic Values, and Scientific Practice.” *Philosophy of Science* 84 (1): 160–80. <https://doi.org/10.1086/688939>.
- Franco, Paul L. 2019. “Speech Act Theory and the Multiple Aims of Science.” *Philosophy of Science* 86 (5): 1005–15. <https://doi.org/10.1086/705452>.
- Fritsch, Matthias. 2023. “Climate Change and Democracy.” In *Handbook of the Philosophy of Climate Change*, edited by Gianfranco Pellegrino and Marcello Di Paola. Springer. https://doi.org/10.1007/978-3-031-07002-0_142.
- Gerovitch, Slava. 1996. “Perestroika of the History of Technology and Science in the USSR: Changes in the Discourse.” *Technology and Culture* 37 (1): 102–34.
- Gerovitch, Slava. 1998. “Writing History in the Present Tense: Cold War-Era Discursive Strategies of Soviet Historians of Science and Technology.” In *Universities and Empire: Money and Politics in the Social Sciences During the Cold War*, edited by Christopher Simpson. The New Press.

- Gerovitch, Slava. 2002. *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*. MIT Press.
- Gerovitch, Slava. 2011. "Why Are We Telling Lies? The Creation of Soviet Space History Myths." *The Russian Review (Stanford)* 70 (3): 460–84. <https://doi.org/10.1111/j.1467-9434.2011.00624.x>.
- Getty, J. Arch. 2008. "Comment: Codes and Confessions." *Slavic Review* 67 (3): 711–5.
- Gill, Graeme. 2011. *Symbols and Legitimacy in Soviet Politics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511791437.001>.
- Gordin, Michael D. 2022. "Lysenkoism." *Encyclopedia of the History of Science*. <https://doi.org/10.34758/d5bq-k368>.
- Graham, Loren R. 1987. *Science, Philosophy, and Human Behaviour in the Soviet Union*. Columbia University Press.
- Greene, Amanda R. 2017. "Legitimacy without Liberalism: A Defense of Max Weber's Standard of Political Legitimacy." *Analyse & Kritik* 39 (2): 295–324. <https://doi.org/10.1515/auk-2017-0017>.
- Hall, Karl. 2008. "The Schooling of Lev Landau: The European Context of Postrevolutionary Soviet Theoretical Physics." *Osiris* 23 (1): 230–59.
- Hilligardt, Hannah. 2023. "Partisan Science and the Democratic Legitimacy Ideal." *Synthese* 202 (5): 135. <https://doi.org/10.1007/s11229-023-04370-5>.
- Hilligardt, Hannah. 2024. "Science as Public Service." *European Journal for Philosophy of Science* 14 (3): 45. <https://doi.org/10.1007/s13194-024-00607-3>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- John, Stephen. 2019. "Science, Truth and Dictatorship: Wishful Thinking or Wishful Speaking?" *Studies in History and Philosophy of Science Part A* 78: 64–72. <https://doi.org/10.1016/j.shpsa.2018.12.003>.
- Josephson, Paul R. 1988. "Science Policy in the Soviet Union, 1917–1927." *Minerva* 26 (3): 342–69. <https://www.jstor.org/stable/41820731>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Kojevnikov, Alexei B. 2002. "The Great War, the Russian Civil War, and the Invention of Big Science." *Science in Context* 15 (2): 239–75. <https://doi.org/10.1017/s0269889702000443>.
- Kojevnikov, Alexei B. 2004. *Stalin's Great Science*. World Scientific Publishing Co. <https://doi.org/10.1142/p307>.
- Kolchinsky, Eduard I. 2014. "Nikolai Vavilov in the years of Stalin's 'Revolution from Above' (1929–1932)." *Centaurus* 56: 330–58. <https://doi.org/10.1111/1600-0498.12059>.
- Kotkin, Stephen. 1995. *Magnetic Mountain: Stalinism as a Civilization*. University of California Press.
- Kotkin, Stephen. 1998. "1991 and the Russian Revolution: Sources, Conceptual Categories, Analytical Frameworks." *The Journal of Modern History* 70 (2): 384–425. <https://doi.org/10.1086/235073>.
- Kotkin, Stephen. 2001. "Really-Existing Revisionism?" *Kritika: Explorations in Russian and Eurasian History* 2 (4): 707–11.
- Krementsov, Nikolai, and William deJong-Lambert. 2017. "'Lysenkoism' Redux: Introduction." In *The Lysenko Controversy as a Global Phenomenon: Genetics and Agriculture in the Soviet Union and Beyond Vol. 1*, edited by William deJong-Lambert and Nikolai Krementsov. Palgrave McMillan. https://doi.org/10.1007/978-3-319-39176-2_1.
- Le Bihan, Soazig. 2024. "How to Not Secure Public Trust in Science: Representative Values Versus Polarization and Marginalization." *Philosophy of Science* 91 (5): 1128–38. <https://doi.org/10.1017/psa.2023.119>.
- Lusk, Greg. 2020. "Political Legitimacy in the Democratic View: The Case of Climate Services." *Philosophy of Science* 87 (5): 991–1002. <https://doi.org/10.1086/710803>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- McKaughan, Daniel J. 2007. "Toward a Richer Vocabulary for Epistemic Attitudes: Mapping the Cognitive Landscape." PhD diss., University of Notre Dame.
- McKaughan, Daniel J., and Kevin C. Elliott. 2013. "Backtracking and the Ethics of Framing: Lessons from Voles and Vasopressin." *Accountability in Research* 20 (3): 206–26. <https://doi.org/10.1080/08989621.2013.788384>.
- Mittiga, Ross. 2022. "Political Legitimacy, Authoritarianism, and Climate Change." *American Political Science Review* 116 (3): 998–1011. <https://doi.org/10.1017/S0003055421001301>.
- Peter, Fabienne. 2023. *The Grounds of Political Legitimacy*. Oxford University Press.

- Pulkkinen, Karoliina. 2025. "Legitimacy Through Outputs: How Science and Technology Can Contribute to Nondemocratic Legitimization." Unpublished manuscript.
- Roll-Hansen, Nils. 2004. *The Lysenko Effect. The Politics of Science*. Humanity Books.
- Roll-Hansen, Nils. 2015. "On the Philosophical Roots of Today's Science Policy: Any Lessons from the 'Lysenko Affair'?" *Studies in East European Thought* 67 (1/2): 91–109. <https://doi.org/10.1007/s11212-015-9231-8>.
- Schroeder, S. Andrew. 2017. "Using Democratic Values in Science: An Objection and (Partial) Response." *Philosophy of Science* 84 (5): 1044–54. <https://doi.org/10.1086/694006>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Schroeder, S. Andrew. 2022. "The Limits of Democratizing Science: When Scientists Should Ignore the Public." *Philosophy of Science* 89 (5): 1034–43. <https://doi.org/10.1017/psa.2022.54>.
- Steffek, Jens. 2015. "The Output Legitimacy of International Organizations and the Global Public Interest." *International Theory* 7 (2): 263–93. <https://doi.org/10.1017/S1752971915000044>.
- Thoma, Johanna. 2024. "Social Science, Policy and Democracy." *Philosophy & Public Affairs* 52 (1): 5–41. <https://doi.org/10.1111/papa.12250>.
- Williams, Bernard. 2005. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*, edited by Geoffrey Hawthorn. Princeton University Press.

23

SCIENCE AND SOCIAL JUSTICE

Heather Douglas

23.1 Introduction

Science is a source of power. This claim should be uncontroversial, whether one thinks science always (eventually) gets at the truth, or, more modestly, is the most reliable source of empirical knowledge available (but is still highly fallible), or is merely a knowledge claim producer that has been granted more cultural authority than is warranted. (My own view is the middle claim.) Regardless of which view of science's knowledge-production capacities one endorses, science is powerful because of its cultural position at the center of empirical knowledge production. Science informs the causal claims we make, the kinds of interventions in the world we attempt, and the categories deemed relevant for both the articulation of public goods and the pursuit of private interest. Even in an age of alternative facts, disputants in the public sphere make appeals to the authority of science. Because science is a source of power, how that power gets distributed is a concern of social justice.

Most centrally, how power directs the efforts of science can be analyzed through the lens of justice. Justice concerns both fair processes of social institutions and fair distributions of social goods. Science is both a social institution and a producer of social goods. Science does not produce knowledge claims at random. The efforts of scientists are directed, as Kitcher noted over 20 years ago, at significant truths (Kitcher 2001, chap. 6). The question is, significant for whom? How research questions get framed, which methodological approaches are deemed best (or good enough), and what is deemed sufficient evidence for a claim are crucial moments of judgment in science, and are (and should be) shot through with social and ethical values (Douglas 2009, 2016; Elliott 2017, 2022). These values are often contested, and the more powerful segments of society can often insist on the primacy of their values for the judgments that direct scientific work. In this way, the existing power structures can recruit and utilize science to reinforce those power structures. In the worst cases, the power of science gets deployed to reinforce or exacerbate unjust inequalities, furthering the interests of the already (unjustly) powerful (Harding 2006; Irzik and Kurtulmus 2024).

For the purposes of this chapter, I will not presume that every inequality—not even every inequality of power—is an injustice. Elected representatives hold more power than the rest of us, but such an inequality of power is not in itself an injustice, but rather needed to justly govern large, complex democracies. Admittedly, we need to ensure that genuine recall power still rests with the electorate and that we can properly assess the actions of our representative (an assessment that requires, minimally, transparency laws and a free press), but the inequality of power inherent in representative

democracies is not in itself unjust. It is an ongoing question precisely when inequalities are just or unjust, and I will not attempt to settle the issue here.¹ For this overview, it suffices that there are clear cases when the imbalance of power and resources within science is unjust. I will use those to illustrate several areas within which we should be concerned about injustice in science.

I describe here four areas of concern for justice in science: (1) Who gets to do science, (2) the use of human subjects in science, (3) research engagement with communities, and (4) setting the research agenda. In each section, I will discuss cases and examples of unjust scientific practice and examples of how science can pursue justice. These four areas of concern are not meant to be a complete topography of science and justice issues.² I present these four areas to show some of the range of concerns that arise and to illustrate unjust and more just practices in science. Additional areas (such as linguistic concerns, ontological concerns, and modeling concerns) can also be described. For the sake of brevity, however, I will attend to these four.

In some areas, there is clear alignment between the epistemic goals of science and the goal of justice. But generally, we will see that there is no automatic or natural alignment between epistemic and justice concerns, which means scientists and philosophers of science must attend to both in their work if science is to serve the cause of justice.

23.2 Access to doing science

The most obvious way in which science has historically been unjust is in the exclusionary way in which it was decided who got to do science. The accounts of sexism and racism in who was included in the ranks of professional scientists are legion, as are the accounts of the erasure of women's and people's of color contributions to science (Harding 1993; Schiebinger 1999). From Marie Curie's difficulties in securing lab space at the beginning of the twentieth century to twenty-first-century female scientists and scientists of color continuing to endure slights and sidelining, science has been a difficult field to pursue for non-white non-male (Harmon 2019; Pickett 2019). Such exclusionary tactics have hampered the development of science (because talented people have been held back), while also being unjust (because the restriction is arbitrary and harms those worse off in society) (Intemann 2009). While one can (regrettably) find those who will still defend or instantiate such exclusions, there are no well-informed defenses of such practices (Giuffrida and Busby 2018; Le Page 2018; Williams 2017).

As many have argued, there are good reasons to ensure that the doors of science are open to all (Intemann 2009; Longino 2002; Merton 1942). At least since Merton's (1942) essay, science has been, in theory, aiming to include anyone who wants to pursue science. It has admittedly not been doing this very well, but progress has been made. We have shifted from cultures of overt bias (in which Robert Millikan could complain about there being too many Jews in physics and explicitly argue for hiring men over women in the lab—see Goodstein (2010, 49)) to cultures where overt bias is no longer acceptable, and we are working to rid science of the microaggressions that still pervade the lab. Perhaps the most hopeful shift is from not just detecting bias and mitigating it post hoc but arguing for microaffirmations that make everyone feel actually welcome. We have ways to fight the biases that still undercut Merton's universalism, and they are being deployed (Estrada, Young, and Nagy 2019; Koch et al. 2022). This is important not just because it is the right thing to do ethically, not just because it will produce better knowledge (Intemann 2009), but also because it is part of making a more just society.

Here, the epistemic, ethical, and political align, and one can use any of these normative bases to argue for the same practical result—an elimination of these structural barriers and full equitable support for all who pursue becoming scientists. This makes the recommendation to open the doors of science, genuinely and in a fully welcoming way, normatively straightforward, even if practically

challenging. Such ease of alignment is not present for the concerns discussed in the next three sections. Furthermore, the exclusionary tactics of white male scientific cultures have not just had an impact on who gets to do science. They have also impacted how and which science gets done.

23.3 The use of human subjects

The next most obvious example of injustice in science concerns the abuse of human subjects. The most egregious examples, unsurprisingly, concern the abuse of society's most vulnerable, and it is these examples that show not just unethical behavior by scientists, but also unjust behavior.

The twentieth century is replete with examples of scientists abusing the most vulnerable people to pursue their research. (It is an abuse of the vulnerable that is a justice concern, not just an ethics concern.) One need not turn to the Nazi doctor atrocities committed in concentration camps to illustrate the problem. There was also the infamous Tuskegee syphilis study, which ran from 1932 to 1972 and followed impoverished African American men with syphilis to see what the course of the disease was when left untreated ([Brandt 1978](#)). Note that there was treatment available in 1932 (Salvarsan) and that even when penicillin became widely available after WWII and was confirmed to be very effective against syphilis, the men in the study remained untreated. They were also deceived about this fact, being told they were being treated (both in person and in advertisements for the study). The names of the men in the study were also distributed to other sources of medical care in the area (even the Army draft board) with the request that the men not receive any treatment for syphilis ([Brandt 1978](#)). The racism and abuse of power in this 40-year study are astonishing. To make a terrible situation even worse, nothing was learned from this study, premised as it was on the racist assumption that black bodies would respond to syphilis differently. If something had been learned, it would not have justified such an egregious abuse of power—not in the slightest. But it is important to note that the study was an epistemic failure as well as a moral and political failure.

Similar racism drove the nutritional experiments among residential school children in Canada ([Mosby 2013](#)). These studies, which took place primarily after WWII, abused indigenous children who were already suffering from the residential school system. This system, set up to impose settler language and culture on indigenous peoples, forcibly removed children from their homes and relocated them to poorly run and poorly resourced residential schools, where the use of their native language and practice of their native culture were forbidden. In this already deeply unjust context, researchers decided to take advantage of these captive and vulnerable children for nutritional research. Children were placed on (or simply continued on) diets that lacked key nutrients to see what happened. No consent for these studies was obtained (from parents or children). Like the Tuskegee study, little of scientific value was learned from these abuses ([Mosby 2013](#)).

These were not isolated incidents. The list of unjust human experimentation conducted in and by democratic societies is long. There is a history of experiments on prisoners (such as the dermatology experiments at Holmesburg prison in the US) ([Hornblum 1998](#)). There is a history of experiments conducted by the military on their own soldiers, including studies premised on racist assumptions that different races might react to chemical weapons differently ([Smith 2008](#)). There were experiments conducted in colonial or imperial contexts of impunity, such as the Guatemala syphilis experiments (which differed from Tuskegee both by the fact that people were deliberately given syphilis but also that genuine treatment was offered) ([Reverby 2011](#)). This is not a comprehensive list.

It should be noted that not all ethically problematic cases of research are necessarily unjust.³ Although justice is considered one of the key principles of medical ethics, it does not get the central attention of beneficence and autonomy and is often only flagged for particular populations (e.g., children, prisoners). If human subjects are targeted for research because of their vulnerability, issues of justice arise. It remains important to attend to issues of justice in human subject research because

we have not left this abuse of the vulnerable behind us, in the pre-IRB, pre-Belmont Report era. Unjust human experimentation still occurs, most notably in the form of “ethics dumping,” in which scientists find populations (often in low- or middle-income countries) that have desired baseline conditions (such as poor access to standard medical care), little power, and lax ethical oversight (Schroeder et al. 2018).

Unlike the issue of access to scientific careers, normative goals come apart when considering human subject research. The goal of justice, and of ethical treatment of human subjects more broadly, often pulls against the goals of knowledge production in particular instances. The epistemic can be in conflict with the ethical and the just, and there is no natural alignment of these normative concerns. It is true that many of the worst abuses of the past (Tuskegee, residential school studies, and many Nazi experiments) produced little knowledge of value, but that is not universal. There were Nazi experiments that did produce valuable knowledge (e.g., their horrific experiments on hypothermia), and contemporary ethics dumping can produce valuable epistemic insights into treatment regimes. What scientists often think will lead them to the best knowledge can run roughshod over principles of ethics or justice—and when this happens, scientists should *not* think that any potential knowledge will make such violations acceptable. It never does, even when valuable knowledge is produced.

What does explicit attention to justice look like in the context of human subject research? Minimally, scientists have to be careful when working with human subjects to be sure that their instincts about what will produce the best data do not lead them to act in ways that are unethical or unjust, that harm their human subjects or take advantage of their vulnerability, for the sake of scientific progress or new knowledge. And thinking that because their subjects are already so badly off, that surely the study’s effects would be an improvement is to reinforce, rather than ameliorate, the power imbalances in the scientist–human subject relationship, particularly in contexts already rife with unjust inequalities. The epistemic does not trump consideration of ethics or justice when grappling with human subjects.

More ideally, we should pay attention to calls for informed consent reform (e.g., moving away from legalistic documents to processes that actually inform participants and having people who are not running the study run the informed consent process). We should ensure that institutional oversight includes substantial numbers of non-scientists and non-academics, so that ethical and justice concerns do not get drowned out in the conversation; token non-scientists are insufficient on such bodies. And we should think about moving away from the language of “human subjects,” which implies a passivity that is difficult to square with the autonomy they are supposed to have, and consider different terms, such as human participants. Because the epistemic and the just can come apart, and pull in opposite directions in many cases, it is crucial that scientists, and those overseeing their work, take seriously the inherent normative tensions and address them explicitly in considering research on human beings.

23.4 Engagement with human communities

Scientists do not have to recruit individual human subjects into research studies to commit or exacerbate injustices. Scientists can (and have) abused human communities in the process of doing their work, even when they are not directly studying living human beings.

For example, for most of the twentieth century, archeologists strove to develop a professional identity and ethic that placed themselves as the authorized interpreters of the human cultural past through artifacts, “as authoritative experts on a resource of great public significance that they are best fitted to document, appraise, and exploit” (Wylie 2002, 234). As a result, archeologists saw archeological sites as a resource about which they should be the ones making decisions. To some indigenous groups, whose sacred sites are among the resources archeologists would like to exploit,

“archaeologists are nothing more than glorified looters” (Wylie 2002, 243). Some indigenous groups do not want their cultural sites preserved; others don’t want their cultural sites locked in stasis for the benefit of archeologists (Wylie 2002, 242). Regardless, most archeologists, for most of the twentieth century, were not interested in sharing decision-making about these resources with indigenous groups.

Since 1990, we have seen passage in the US of the Native American Graves Protection and Repatriation Act (1990) and the World Archaeology Congress’s formal recognition of the need for archeologists “to seek formal consent from, and to actively consult and collaborate with, any indigenous groups whose heritage is the subject of archaeological investigation” (Wylie 2002, 244). Such requirements have changed the efforts of many archeologists, who now embrace the requirement for active and mutually respectful partnerships with indigenous groups in their work. (See also Wylie 2014.) The idea that the knowledge which might be gleaned for humanity writ large means that it is acceptable to run roughshod over the concerns and interests of already oppressed people who have a clear and strong interest in the resource archeologists want to exploit has largely been replaced with a recognition that such a pattern of scientific practice is unacceptable and unjust.

Other scientific fields, however, have not yet had such a realization. In some paleogenomic research attempting to determine the paths of ancient human migration, for example, we again have a disrespectful and fundamentally extractive relationship between scientists and particular communities⁴ (Havstad 2022; Lewis-Kraus 2019). But human communities are not mere epistemic resources for scientists and should never be thought of as such by scientists. Particularly when those human communities have borne the brunt of a history of exploitation, colonialism, and oppression, for scientists to continue those patterns is to commit an injustice. The pursuit of knowledge does not in itself justify such actions. Scientists thus must be alert to the social and political history of the human communities with which they interact and must be on guard against repeating exploitative patterns.

How should scientists do this? When a scientist works with a community, it should not be a purely extractive relationship. The community is not there just to make the science better. The scientist should work within a sense of partnership and collaboration with the community, and obligations to that community should follow from this partnership (Simons, Martindale, and Wylie 2021). This kind of partnership can be difficult to build. It takes genuine commitment to understand and engage with the community and to do so as a long-term effort. This can mean coming to the partnership with respect for, and a readiness to collaborate with, the existing knowledge systems of the community (Whyte 2013). Whether in partnering with fenceline, impoverished, and/or indigenous communities, such collaborations can leverage the power of science to help sharpen the detection of, and the amelioration of, injustices, while also furthering the epistemic goals of science and of the partner communities (Krimsky 2003, chap. 11; Ottinger 2013, 2018).

23.5 The research agenda

One could see the concerns of the previous two sections as being about value judgments in methodological issues, but not about what knowledge we are trying to pursue. One could see such concerns as showing that while methodological restrictions are clearly necessary—because the pursuit of knowledge is not such an overwhelming good that all other concerns pale before it—our underlying desire for knowledge and our assessments of what is interesting, important, or significant to know remain unchanged. We might need to alter our methodologies in order to get at the knowledge we seek, to come at it from other ways, even to accept lower standards of evidential strength (e.g., using animal subjects instead of human subjects for toxicological studies), but we would still strive for the same answers to the same questions.

Such an analysis would fail to acknowledge the importance of the scientific research agenda for issues of justice and the way in which research agenda setting, irrespective of methodological concerns, can reflect patterns of power and injustice. How *what* we decide to know, not just *how* we decide to pursue knowledge, is shaped by power and can be itself a source of injustice.

A prominent example of problematic power interests shaping science, that of the influence commercial interests exert over the distribution of research effort, is already well-known among philosophers of science, particularly in medical research (Holman and Elliott 2018; Radder 2019). Attention to some diseases or lines of treatment swamps attention to others, mostly because of who gets the diseases and whether the potential treatment could make money (Reiss 2010; Reiss and Kitcher 2009). Sheldon Krinsky has called this turn in medical research a pursuit of science in the private interest and argues for a return to science in the public interest (Krinsky 2003). Science in the private interest is paid for and directed by the concerns of wealthy corporations that can afford to pay for expensive scientific research. Philosophers have described cases in which diseases are manufactured so that drugs can be approved and sold for them (Brown 2008; Holman and Geislar 2018), in which the medical research is bent or distorted to show that patentable drugs that don't help people do help people (McGarity and Wagner 2008), and in which medical interventions actually do far more harm than good, yet still evidence is produced to get regulatory approval and widespread sales (Biddle 2007; Holman 2019). These cases often involve epistemic, ethical, and justice failings together, but in complex ways. For example, research in patentable drugs for Western consumers, as opposed to nonpatentable exercise or diet changes, might be epistemically sound as far as the questions being asked. But because epistemic effort and resources are poured into patentable drugs, other questions get ignored (Brown 2008). That can distort what we know as well as how that knowledge gets deployed to worsen existing injustices.

This is not just a concern in biomedical research. It has also been a concern in agribiotech research for over 30 years, since Martha Crouch's thought-provoking essays. Crouch raised concerns over the role of scientists in helping industrial agriculture worsen environmental conditions and harm indigenous groups (Crouch 1990). She argued that more attention was needed by scientists to the broader socio-economic systems in which their research, even apparently "basic" research, would be deployed, and to choose their research agendas accordingly (Crouch 1991). It is also a concern in the geosciences because of their role in extractive industries and colonial expansion (Ohnesorge and Watkins 2024). These concerns become more pronounced with the recognition that private dollars now outpace public dollars for science funding across most of the OECD countries.

In this milieu, all scientists should actively worry whether their research in fact serves the public good, and this concern must take into account the distribution of the societal benefits and harms of research results, a concern over justice. Scientists should ask whether their research will more likely serve the interests of the few, thus reinforcing the already concerning inequalities that saturate our world. For example:

- Should researchers improve facial recognition software when the primary use of such software is not to make it easier to unlock your phone, but to increase the ability of authoritarian regimes to track and monitor their citizens? Even if such technology is used only in democratic and open societies, there are ongoing concerns about the biases in such software (the problems with inaccuracies for people of color) and the way in which the use of the technology will exacerbate existing injustices in our law enforcement systems (Hill 2023).⁵
- Should anyone pursue human immortality (not just morbidity reduction), whether such immortality would be *in vivo* or *in silico*, given that, as a practical matter, such immortality would likely be reserved for the very wealthy and would thus make worse existing and already problematic inequalities?⁶

- Should any scientist ask research questions about cognitive differences among genders or races, given the history of that research (weak methods producing racist results, strong methods erasing them)? There is also the weak practical import of this work, as differences within those groups have always been much larger than differences between them, and thus, we should never use racial or gender identity as any kind of proxy for assessment of a person's abilities. If all such research just feeds the beast of ongoing racism and sexism, there seems to be no good reason to pursue it (Kourany 2016).

It is hard to see how the pursuit of these research agendas does anything for the vast majority of humanity other than make their lives more difficult or precarious. These are just three examples from different fields of study. There are certainly more. Irrespective of methodological concerns or access concerns, there are justice concerns that can and should be raised regarding the shaping of the scientific research agenda. What counts as knowledge worth pursuing, and the resources devoted to such agendas, is an essential aspect of justice in science. It should be a general worry among scientists whether their research will serve the interests solely of those already in power—reinforcing power imbalances or strengthening the hand of autocratic governments—or whether their work is genuinely in the interest of *all* of humanity, not just the elite few.

Note as well, as with the discussion of human subjects and human communities, the epistemic is neither an adequate cover for practices that produce or worsen injustice, nor is there a natural alignment between the epistemic, the ethical, and the just. One can be pursuing knowledge, even for its own sake, and making social injustice worse. This is what Martha Crouch so painfully discovered in her own research. Because the epistemic and the just do not naturally align, scientists should attend to justice concerns in addition to epistemic concerns.

How might this impact the pursuit of science? In some cases, the more just approach might mean giving up the global intervention (scope) in favor of the local one (see also Nieves Delgado, Baedke and Fischer 2026 in this volume for global vs. local research considerations). For example, the pursuit of agroecological solutions necessitates attention to local conditions for crop production, rather than attempting to create the genetic modifications that allow a crop to grow across a range of conditions (something of great interest to corporate agri-business) (Lacey 2005). The global solution might look more attractive, but the scientist needs to attend to the existing power relations, such as the impact of intellectual property rights on the genetic material of the crops being developed, to assess whether the pursuit of a particular agenda will likely worsen rather than ameliorate injustice (Hicks 2017). In other cases, a more global solution—such as a vaccine for a neglected or emerging disease that can be deployed inexpensively across the globe—might be exactly what will help the most.

In general, scientists should ensure their research agendas are not obviously unjust (which would make the world worse in clearly foreseeable ways), and ideally should aim for improving societal justice (Douglas 2024). There are many ways to accomplish this. For example, scientists can collaborate with local communities (as discussed above). They can participate in research movements within science that attempt to address injustices (Koskinen and Rolin 2019). They can aim to innovate with express attention to inequalities (Parthasarathy 2022). Regardless of how, attending to aspects of justice in the shaping of the research agenda is needed.

23.6 Conclusions: Attending to power

The arc of history can bend toward justice, but only if we make it so. We need to take seriously the idea that science is a source of power, and thus that it can be wielded either justly or unjustly. Scientists (and academics more generally) need to attend to not just their obligations to pursue truth and

institutionalized ethical constraints, but also their obligations to pursue justice, as part of attending to the impact of their work on society.

We want science that pursues true (or at least reliable) knowledge, that is ethical, and that is just.⁷ These goals, we should acknowledge, do not always pull together. What is good for knowledge production can be in alignment with what is good for justice, *or not*. With respect to who gets to do science (issues of access to scientific careers), epistemic aims are usually furthered by pursuing just access. But, in general, there is no guarantee these goals will always align.

Are there cases when the desire for knowledge should override justice concerns? We should be wary of claims that epistemic good should override considerations of justice. First, many scientific research efforts do not pan out; promises of breakthroughs and grand insights are often hollow. Strong instincts that a research project will be hugely beneficial, if only we can just use this one population or ignore this one problem of injustice, should be met with potent skepticism. The examples of Tuskegee and the residential school nutritional studies, which produced no valuable knowledge, should serve as stark reminders of the depths of epistemic and moral failure to which such reasoning can descend. Given the resource constraints and problems facing humanity, we probably need science to be aiming at all three normative goals (the epistemic, the ethical, and the just) all the time if we are going to make it to the other side of the twenty-first century (Turnhout 2024).

How can philosophers of science help make science an instrument of justice? Part of the answer is in helping to reveal how science has been part of committing injustices (and is continuing to be part of such practices). Part of the answer is in describing what science in the service of justice looks like. I aimed to provide a bit of both in this chapter. Finally, descriptions of what kinds of policies would help scientists do work that serves the ends of justice are needed. This is beyond my scope here, but the chapters in [Part IV](#) of this handbook begin turning to these institutional questions.

Ideally, the values that shape crucial scientific judgments (about what is significant to know, about which kinds of errors are acceptable, about what categories or causes are important) would be carefully developed and well-reasoned value judgments shared by all (Brown 2020). But the history of science reveals that this has not been the case, it is not the case now, and it needs to be a matter of conscious effort to make it the case going forward. Existing power will try to bend science (as a source of power) toward its own interests. If there are injustices built into existing power (as so often there are), that bends science toward injustice. Making science just will require explicit and concerted effort, an effort in which philosophers of science can play a crucial part.

This is an admittedly brief survey of some ways that science can assist the aim of social justice. What can be gained by shifting to a world where all people have the possibility of becoming scientists, where non-scientists (both as individuals and as communities) are looked upon as partners of scientists, and where research agendas are shaped by genuine communal concerns? We can see the promise of such an approach in the example of Charles Menzies, a First Nation archeologist who discovered robust archeological evidence that First Nations in British Columbia had been harvesting and eating abalone for thousands of years (Menzies 2015). Why was this important? Canadian fisheries managers working for the government had presumed that abalone was not abundant until sea otters were hunted to near extinction about a century ago, and so there was no traditional right of First Nations to harvest abalone. Menzies, working in collaboration with First Nations who had strong oral histories of abalone harvesting and processing extending well back into the nineteenth century, found the archeological sites that proved the Canadian government wrong and the First Nations' oral history right. This example shows how we can have better knowledge (reaching the epistemic goal), generated in just ways (in genuine collaboration and partnership with a community), produced by someone who would have been excluded from science just decades ago, and that helps redress a historical and contemporary injustice. This is what science can be.

Notes

- 1 Aristotle thought that benefits should be commensurate with burdens. Rawls thought that inequalities should benefit everyone, especially the least well-off. These are just two accounts of when an inequality may be just, and they don't touch on issues of historical injustice.
- 2 I also will not be using the testimonial or hermeneutic injustice framings, although one could construe the exclusion of some people from science as a form of testimonial injustice and some research agenda-setting issues as ones of hermeneutic injustice. See Fricker (2007) and the chapter by Gutiérrez Valderrama in this handbook (Gutiérrez Valderrama 2026) for more discussion of these issues.
- 3 For example, highly educated patients who suffer from therapeutic misconception are an ethical issue, but not clearly a justice issue.
- 4 "As one ancient-DNA researcher in Turkey put it to me, 'Certain geneticists see the rest of world as the 19th-century colonialists saw Africa—as raw-material opportunities and nothing else'" (Lewis-Kraus 2019).
- 5 See the discussion on these issues at forums such as <https://www.aaas.org/events/responsible-ai-facial-recognition-and-scientific-responsibility>.
- 6 <https://intelligencesquaredus.org/debates/lifespans-are-long-enough>.
- 7 Janet Kourany's Socially Responsible Science ideal articulates a joint satisfaction ideal, demanding that science be both empirically successful and socially successful (Kourany 2008, 68). The view I am articulating here emphasizes issues of justice in addition to issues of ethics for science to be considered socially successful.

References

- Biddle, Justin. 2007. "Lessons from the Vioxx Debacle: What the Privatization of Science Can Teach Us about Social Epistemology." *Social Epistemology* 21 (1): 21–39. <https://doi.org/10.1080/02691720601125472>.
- Brandt, Allan M. 1978. "Racism and Research: The Case of the Tuskegee Syphilis Study." *Hastings Center Report*: 21–9. <https://doi.org/10.2307/3561468>.
- Brown, James Robert. 2008. "The Community of Science®." In *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, edited by Martin Carrier, Don Howard, and Janet Kourany. University of Pittsburgh Press.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Crouch, Martha. 1990. "Debating the Responsibilities of Plant Scientists in the Decade of the Environment." *The Plant Cell* 2 (4): 275–7. <https://doi.org/10.1105/tpc.2.4.275>.
- Crouch, Martha. 1991. "The Very Structure of Scientific Research Mitigates against Developing Products to Help the Environment, the Poor, and the Hungry." *Journal of Agricultural and Environmental Ethics* 4 (2): 151–8. <https://doi.org/10.1007/BF01980311>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2016. "Values in Science." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.
- Douglas, Heather. 2024. "Structuring Institutions for Responsible and Accountable Science." *Philosophy of Science* 91 (5): 1149–58. <https://doi.org/10.1017/psa.2023.128>.
- Elliott, Kevin. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Estrada, Mica, Gerald R. Young, Jill Nagy, et al. 2019. "The Influence of Microaffirmations on Undergraduate Persistence in Science Career Pathways." *CBE—Life Sciences Education* 18 (3): ar40. <https://doi.org/10.1187/cbe.19-01-0012>.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Giuffrida, Angela, and Mattha Busby. 2018. "‘Physics Was Built by Men’: CERN Suspends Scientist over Remarks." *The Guardian*, October 1. <https://www.theguardian.com/science/2018/oct/01/physics-was-built-by-men-cern-scientist-alessandro-strumia-remark-sparks-fury>.
- Goodstein, David. 2010. *On Fact and Fraud: Cautionary Tales from the Front Lines of Science*. Princeton University Press.
- Gutiérrez Valderrama, Juliana. 2026. "Values in Global Science and the Relevance of Geographic Diversity." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Harding, Sandra. 1993. *The "Racial" Economy of Science: Toward a Democratic Future*. Indiana University Press.

- Harding, Sandra. 2006. *Science and Social Inequality*. University of Illinois Press.
- Harmon, Amy. 2019. "For a Black Mathematician, What It's Like to Be the 'Only One'." *The New York Times*, February 18. <https://www.nytimes.com/2019/02/18/us/edray-goins-black-mathematicians.html>.
- Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Hicks, Daniel J. 2017. "Genetically Modified Crops, Inclusion, and Democracy." *Perspectives on Science* 25 (4): 488–520. https://doi.org/10.1162/POSC_a_00251.
- Hill, Kashmir. 2023. *Your Face Belongs to Us: A Secretive Startup's Quest to End Privacy as We Know It*. Random House.
- Holman, Bennett. 2019. "Philosophers on Drugs." *Synthese* 196: 4363–90. <https://doi.org/10.1007/s11229-017-1642-2>.
- Holman, Bennett, and Kevin C. Elliott. 2018. "The Promise and Perils of Industry-Funded Science." *Philosophy Compass* 13 (11): e12544. <https://doi.org/10.1111/phc3.12544>.
- Holman, Bennett, and Sally Geislar. 2018. "Sex Drugs and Corporate Ventriloquism: How to Evaluate Science Policies Intended to Manage Industry-Funded Bias." *Philosophy of Science* 85 (5): 869–81. <https://doi.org/10.1086/699713>.
- Hornblum, Allen M. 1998. *Acres of Skin: Human Experiments at Holmesburg Prison*. Routledge.
- Intemann, Kristen. 2009. "Why Diversity Matters: Understanding and Applying the Diversity Component of the National Science Foundation's Broader Impacts Criterion." *Social Epistemology* 23 (3–4): 249–66. <https://doi.org/10.1080/02691720903364134>.
- Irzik, Gürol, and Faik Kurtulmus. 2024. "Distributive Epistemic Justice in Science." *The British Journal for the Philosophy of Science* 75 (2): 325–45. <https://doi.org/10.1086/715351>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Koch, Julie M., Douglas Knutson, LaTrice Loche, Rogers W. Loche III, Hang-Shim Lee, and Dillon J. Federici. 2022. "A Qualitative Inquiry of Microaffirmation Experiences among Culturally Diverse Graduate Students." *Current Psychology* 41: 2883–95. <https://doi.org/10.1007/s12144-020-00811-3>.
- Koskinen, Inkeri, and Kristina Rolin. 2019. "Scientific/Intellectual Movements Remediating Epistemic Injustice: The Case of Indigenous Studies." *Philosophy of Science* 86 (50): 1052–63. <https://doi.org/10.1086/705522>.
- Kourany, Janet. 2008. *Philosophy of Science after Feminism*. Oxford University Press.
- Kourany, Janet. 2016. "Should Some Knowledge Be Forbidden? The Case of Cognitive Differences Research." *Philosophy of Science* 83 (5): 779–90. <https://doi.org/10.1086/687863>.
- Krimsky, Sheldon. 2003. *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* Rowman & Littlefield.
- Lacey, Hugh. 2005. *Values and Objectivity in Science: The Current Controversy about Transgenic Crops*. Lexington.
- Le Page, Michael. 2018. "Physicist Sparks Gender Row after Claiming Women Are Worse at Physics." *New Scientist*, October 1. <https://www.newscientist.com/article/2181113-physicist-sparks-gender-row-after-claiming-women-are-worse-at-physics/>.
- Lewis-Kraus, Gideon. 2019. "Is Ancient DNA Research Revealing New Truths—Or Falling into Old Traps." *The New York Times Magazine*, January 17. <https://www.nytimes.com/2019/01/17/magazine/ancient-dna-paleogenomics.html>.
- Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.
- McGarity, Thomas O., and Wendy E. Wagner. 2008. *Bending Science: How Special Interests Corrupt Public Health Research*. Harvard University Press.
- Menzies, Charles R. 2015. "Revisiting 'Dm Sibilhaa'nm da Laxyuubm Gitxaala (Picking Abalone in Gitxaala Territory)': Vindication, Appropriation, and Archaeology." *BC Studies: The British Columbian Quarterly* 187: 129–53. <https://doi.org/10.14288/bcs.v0i187.187220>.
- Merton, Robert K. 1942. "A Note on Science and Democracy." *Journal of Legal and Political Sociology* 1: 115–26.
- Mosby, Ian. 2013. "Administering Colonial Science: Nutrition Research and Human Biomedical Experimentation in Aboriginal Communities and Residential Schools, 1942–1952." *Histoire Sociale/Social History* 46 (1): 145–72. <https://doi.org/10.1353/his.2013.0015>.
- Nieves Delgado, Abigail, Jan Baedke, and Ann-Christin Fischer. 2026. "The Limits of Diversity in Science: The Case of Human Microbiome Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Ohnesorge, Miguel, and Aja Watkins. 2024. "What Is Philosophy of the Geosciences?" *Philosophy Compass* 19 (2): e12962. <https://doi.org/10.1111/phc3.12962>.

- Ottinger, Gwen. 2013. *Refining Expertise*. New York University Press.
- Ottinger, Gwen. 2018. "Making the Most of Air Monitoring." https://storage.googleapis.com/publiclab-production/public/system/images/photos/000/026/389/original/Making_the_Most_of_Monitoring_Report.pdf. Accessed July 8, 2025.
- Parthasarathy, Shobita. 2022. "Innovation as a Force for Equity." *Issues in Science and Technology* 38 (2): 30–6. <https://www.jstor.org/stable/27161898>.
- Pickett, Mallory. 2019. "I Want What My Male Colleague Has, and That Will Cost a Few Million Dollars." *The New York Times*, April 18. <https://www.nytimes.com/2019/04/18/magazine/salk-institute-discrimination-science.html>.
- Radder, Hans. 2019. *From Commodification to the Common Good: Reconstructing Science, Technology, and Society*. University of Pittsburgh Press.
- Reiss, Julian. 2010. "In Favour of a Millian Proposal to Reform Biomedical Research." *Synthese* 177 (3): 427–47. <https://doi.org/10.1007/s11229-010-9790-7>.
- Reiss, Julian, and Philip Kitcher. 2009. "Biomedical Research, Neglected Diseases, and Well-Ordered Science." *THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia* 24 (3): 263–82. <https://www.redalyc.org/articulo.oa?id=339730810002>.
- Reverby, Susan. 2011. "'Normal Exposure' and Inoculation Syphilis: A PHS 'Tuskegee' Doctor in Guatemala, 1946–1948." *The Journal of Policy History* 23 (1): 6–28. <https://doi.org/10.1017/S0898030610000291>.
- Schiebinger, Londa. 1999. *Has Feminism Changed Science?* Harvard University Press.
- Schroeder, Doris, Julie Cook Lucas, François Hirsch, Solveig Fenet, and Vasantha Muthuswamy. 2018. *Ethics Dumping: Case Studies from North-South Research Collaborations*. Springer.
- Simons, Eric, Andrew Martindale, and Alison Wylie. 2021. "Bearing Witness: What Can Archaeology Contribute in an Indian Residential School Context?" In *Working with and for Ancestors: Collaboration in the Care and Study of Ancestral Remains*, edited by Chelsea H. Meloche, Laure Spake, and Katherine L. Nichols. Routledge.
- Smith, Susan L. 2008. "Mustard Gas and American Race-Based Human Experimentation in World War II." *Journal of Law, Medicine & Ethics* 36 (3): 517–21. <https://doi.org/10.1111/j.1748-720X.2008.299.x>.
- Turnhout, Esther. 2024. "A Better Knowledge Is Possible: Transforming Environmental Science for Justice and Pluralism." *Environmental Science & Policy* 155: 103729. <https://doi.org/10.1016/j.envsci.2024.103729>.
- Whyte, Kyle Powys. 2013. "On the Role of Traditional Ecological Knowledge as a Collaborative Concept: A Philosophical Study." *Ecological Processes* 2 (1): 7. <https://doi.org/10.1186/2192-1709-2-7>.
- Williams, Lara. 2017. "Memo to All Tech Bros: Sexism, Not Biology, Holds Women Back." *New Scientist*, August 8. <https://www.newscientist.com/article/2143285-memo-to-all-tech-bros-sexism-not-biology-holds-women-back/>.
- Wylie, Alison. 2002. *Thinking from Things: Essays in the Philosophy of Archaeology*. University of California Press.
- Wylie, Alison. 2014. "Community-Based Collaborative Archaeology." In *Philosophy of Social Science: A New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi. Oxford University Press.

ALGORITHMIC ABOLITIONISM AND THE RACIAL ALGORITHM

Alexander Williams Tolbert

24.1 Introduction

The current debate over fairness in machine learning is not just a technical question but an example of a broader philosophical issue regarding the role of non-epistemic values in science (Douglas 2009). The contemporary algorithmic fairness literature asks when the use of machine learning algorithms in socially consequential settings is morally justified rather than only accurate. The literature contains a large taxonomy of fairness definitions and mitigation strategies, grounded in formal analysis. However, despite this, algorithmic systems continue to reproduce racial disparities across several domains, such as healthcare, credit, employment, and criminal justice.

My central claim in this chapter is that algorithmic fairness, as currently theorized, systematically under-theorizes race, abstracting from non-ideal conditions of racial injustice. Given that algorithmic decisions disproportionately harm Black populations, any adequate account of fairness must explain why these groups in particular bear the heaviest burdens. I introduce the concept of the Racial Algorithm to capture these dynamics. The Racial Algorithm is a normative explanatory model of how racialized causal mechanisms, structured by what Mills (1997) calls the “Racial Contract,” produce the observable data distributions on which ML systems are trained, such that the models fitted to those distributions can learn to reproduce the statistical patterns generated by those mechanisms—and how the resulting algorithmic decisions, also shaped by inherited worldviews and institutional practices under the Racial Contract, can serve to lock in and reaffirm racial hierarchies over time. More generally, I argue that the mainstream artificial intelligence (AI) fairness literature reflects a detachment from the conceptual resources and normative orientations developed within Black philosophy.

The chapter proceeds as follows. Section 24.2 reconstructs the incrementalist paradigm that dominates the technical and philosophical literature on algorithmic fairness. Section 24.3 develops the concept of the Racial Algorithm, reframing algorithmic bias structurally. Section 24.4 situates the issue of algorithmic fairness within radical, insurrectionist traditions in ethics. I highlight the tension between reformist approaches, which seek to improve existing systems, and abolitionist approaches, which question whether certain systems should exist at all. I conclude in Section 24.5 by discussing broader debates about the legitimacy of science in democratic societies and conclude by gesturing at a non-ideal theory of algorithmic governance that includes both procedural safeguards and substantive norms.

24.2 Incrementalism: Fairness metrics and modular reform

The debate over algorithmic fairness gained traction with ProPublica's 2016 analysis of COMPAS, which reported that Black defendants who did not reoffend were nearly twice as likely to be labeled "high risk" as comparable to white defendants, even though recidivism rates conditional on risk scores were similar across groups (Angwin et al. 2016). This case highlighted the broader issue of what it would mean, in statistical terms, for an algorithm to be "fair." Further, theoretical work showed that statistical fairness conditions such as predictive parity and equalized odds cannot, in general, be satisfied simultaneously when base rates differ, absent perfect prediction (Kleinberg, Mullainathan, and Raghavan 2016).

In their 2020 survey of the then-nascent AI fairness literature, Chouldechova and Roth observed that "with a few exceptions, the vast majority of work to date on fairness in machine learning has focused on the task of batch classification," attending primarily to "two main families of definitions: statistical notions of fairness and individual notions of fairness" (84). They emphasize that "most of the literature on fair classification focuses on statistical definitions of fairness," where parity of some statistical measure is required across protected demographic groups. Popular measures include "raw positive classification rate... false positive and false negative rates... and positive predictive value" (Chouldechova and Roth 2020, 84). This early orientation toward definitional disputes and technical criteria has proven remarkably durable. Hellman's (2025) Stanford Encyclopedia of Philosophy (SEP) entry on algorithmic fairness similarly frames the terrain in terms of definitional disputes—she maps the field in terms of whether fairness is "comparative" (e.g., like cases should be treated alike) or "non-comparative" (e.g., unfairness as inaccuracy, inscrutability, or arbitrariness) and adds a focused discussion of proxies. As Hellman herself notes, the algorithmic fairness literature is enormous; however, because the SEP is regularly updated and widely regarded as authoritative, we can take Hellman's article to represent one authoritative snapshot of the current state of the field in algorithmic fairness.

As exemplified by the COMPAS controversy, algorithmic systems have consistently overburdened Black populations—including, and in the US context, centrally, descendants of Africans enslaved in the United States. This pattern is not incidental. Rather, it reflects the persistence of a racial order that traces back to slavery, Jim Crow, and continuing regimes of racialized control (McPherson 2024). As McPherson insists, the non-exclusionary "moral urgency and priority, in the United States, of the 'racial' justice claims of Africa-identified descendants of American slavery" must remain central (2024, 144), for only by addressing these claims can we hope to "finally dismantle color caste in the United States" (2024, 170). To say that a certain justice claim has "moral urgency and priority" (in a particular social-political context) is not to claim that it must always be satisfied prior to, or even along with, any generally equal or lesser "moral priority" claim, regardless of immediate urgency and background circumstances. Nonetheless, situating algorithmic harms within this framework makes salient why the burdens borne by Black Americans are not just another statistical fairness violation but the contemporary reflection of a long-standing racial order.

Thus, given that the harms of algorithmic decision-making fall most heavily on Black populations, and centrally on the group already positioned at the bottom of the racial order—descendants of American slavery—it matters how the fairness literature conceptualizes race and what kinds of explanations it provides. When read together, Chouldechova and Roth (2020) and Hellman (2025) illustrate the dominant orientation of the algorithmic fairness literature toward race. Chouldechova and Roth (2020) attend to modular interventions and technical criteria for fairness. Hellman (2025), even while distinguishing comparative from non-comparative notions of fairness, likewise attends to definitional disputes and conceptual refinements. Race is not absent from these accounts, but where it does appear, it is largely perfunctory; it appears in both papers as an example of a protected group, and in Hellman (2025) as part of a short reflection on the nature of social groups more generally.

What is absent across both texts is a normative explanatory model of race—one that explains why algorithms consistently burden African Americans descended from slavery by understanding how racialized causal mechanisms produce data, models, and ultimately decisions. The problem is not only that race is absent but that the literature’s sustained focus on definitions and technical fixes makes clear that structural accounts of racial injustice are peripheral to its concerns.

Thus, the current AI fairness literature reflects what might be called the incrementalist paradigm, which treats bias as a problem that can be addressed using modular technical fixes, rather than by altering institutional contexts or underlying causal mechanisms. This approach is largely consistent with fairness metrics commonly used alongside bias-mitigation techniques. Metrics such as demographic parity, equalized odds, and calibration provide fairness criteria. Bias mitigation methods intervene at different stages of the pipeline—pre-processing training data, in-processing constraints on optimization, or post-processing outputs that may then be further audited. Across these strategies, the underlying assumption remains constant: That bias inheres in algorithms themselves and can be engineered away.

However, a growing body of evidence shows that the incrementalist approach has not resolved the underlying normative problems. Healthcare algorithms can underrate Black patients’ needs because cost proxies reflect unequal access to and utilization of care (Obermeyer et al. 2019). Criminal justice risk tools can produce racial disparities by disproportionately incorporating race-correlated factors such as employment, education, and social environment (Angwin et al. 2016). Machine learning credit models can disproportionately penalize Black and Hispanic borrowers (Fuster et al. 2022). In each case, the source of bias is upstream, in the data-generating structures themselves. This suggests that modular fairness interventions operating downstream on model outputs address symptoms rather than causes. As a result, racial stratification tends to be reproduced rather than remedied.

To be clear, incrementalism can produce some benefits. Minimizing groupwise error rates, for instance, can reduce individual adverse outcomes in contexts where misclassification causes direct harm. For example, suppose we are designing a tool that takes an MRI scan of the liver as input and determines whether it shows a tumor. Both false positives and false negatives have direct harmful consequences for the individual: Needless surgery in the case of a false positive, or an undiagnosed cancer in the case of a false negative. In this setting, error rates for a population correspond directly to the harms it suffers. If the error rate is higher for patients with darker skin, that disparity is unjust, and technical fixes that reduce it are both feasible and necessary. Presumably, everyone would want their error rate to be lower; as a result, accuracy here seems straightforwardly relevant and desirable.¹ However, not all domains resemble medical imaging. Some tasks are inherently impermissible, such that accuracy is irrelevant. For example, predicting abortion risk for punitive enforcement cannot be legitimized by high group-wise accuracy. Other fairness targets are systematically misconceptualized from the outset: For example, treating arrest as a proxy for criminality confuses institutional behavior with individual conduct. In these contexts, downstream technical fixes cannot redeem the enterprise. Thus, the incrementalist paradigm represents a necessary but insufficient response. It is necessary because in some cases, accuracy and parity are direct measures of adverse outcomes, and technical fixes can reduce unjust burdens. It is insufficient because it abstracts away from the historical and institutional structures that generate the data, define the targets, and constrain the uses to which algorithmic decisions are made. Without such a structural account, incrementalist reforms risk producing the illusion of progress while leaving underlying causal mechanisms untouched, thereby mistaking marginal technical improvements for genuine progress.

24.3 From the racial contract to the Racial Algorithm

In the technical and philosophical literature on AI ethics, very little sustained attention is given to the materialities of race and racism’s effects on algorithmic design and data-driven decisions. Mills diagnoses a similar omission in political philosophy. Despite being “supposedly dedicated to the

elucidation of justice,” political theory has for decades made “next to no mention of the centrality of racial in-justice to the ‘basic structure’ of the United States” (Mills 2007, 33). Importantly, Mills observes that “this idealization seems not to be limited to normative theory, but to extend to factual matters as well, including the features of actual society” (2020, 17). When race is acknowledged at all, its treatment is perfunctory, slotted into ideal-theory assumptions that ignore “crucial facts”: Namely, the history of slavery, expropriation, and segregation in the United States (Mills 1994, 226).

This abstraction away from racial realities marginalizes the very populations most affected by injustice and results in what Mills calls a “closed circuit of epistemic authority that reproduces white delusions” (Mills 2007, 34). In his Tanner Lecture, Mills (2020) makes the point with reference to Rawls:

In addition, I pointed out in various critical essays over the years that Rawls’s idealizations seemed not to be limited to normative theory, but to extend to factual matters as well, including the features of actual societies, in a way that ignored the racist record both of the United States in particular and the West more generally. (For example, nowhere in his work does he mention the expropriation and genocide of Native Americans, postbellum Jim Crow segregation, and the historically white-supremacist nature of the United States, nor the Atlantic slave trade or European colonialism and imperialism in the non-European world.) So, this seemed to me to be evidence of the ‘whiteness’ of Rawls (and Rawlsianism also). (16–7)

Mills later elaborates that Rawls was not wholly unaware of race. He cites passages from *A Theory of Justice*, *Political Liberalism*, and *Justice as Fairness* that gesture toward civil disobedience, the recognition of new groups, and “principles for meeting injustice.” However, none of these remarks is ever developed into a substantive treatment of racial injustice. This leads Mills to distinguish between a society “with racism” and a “racist society” (Mills 2020, 26). The former reflects private prejudice and superficial institutional effects, while the latter describes a polity where racism fundamentally shapes the institutions of the basic structure. On Mills’s view, Rawls consistently failed to theorize the United States as a racist society. Hence, his sharp conclusion:

And consider, in the end, as a related closing rebuttal, how obviously implausible—indeed absurd, and in addition, some might say, insulting—it is to think that a Rawls who had lived through the racial tumult of the 1960s, and mentions Martin Luther King Jr. in his later work, would be so indifferent to racial injustice as to devise a theory excluding its correction.

(Mills 2020, 26)

The structural same irony that Mills identifies in liberal political philosophy is repeated in the AI fairness literature. To generate normative urgency for algorithmic fairness, Chouldechova and Roth (2020) and Hellman (2025) take as their central case examples either predictive policing or recidivism tools such as COMPAS, which disproportionately affect Black communities. However, even in the aftermath of George Floyd and the Black Lives Matter movement, neither account places race at the center of analysis. In particular, they fail to center Black people, and in the US context, specifically the descendants of slavery, as the structural category of analysis. Just as Mills treats the traditions’ omissions as evidence of a deeper structural orientation in political philosophy, so too should we treat the fairness literature’s omissions as evidence of its orientation: Inattentive to race and committed instead to a gradualist framework. Its reliance on definitional refinements, modular bias-mitigation techniques, and philosophical models rests on the assumption that these interventions perform meaningful work—even while leaving the structural realities of racial subordination untouched.

The fairness literature, then, does two things. First, it leaves precisely this structural gap: It does not develop an explanatory model of how racial subordination is produced and sustained. Second,

it proceeds within this gradualist, tacitly incrementalist orientation, treating piecemeal refinements and fixes as if they were adequate responses in the absence of a structural theory. What is missing, therefore, is a structural account that does not simply register disparate outcomes but explains them and does so in relation to those who bear non-exclusive moral priority in this domain—most centrally Black people, and in the United States, the descendants of slavery.

I suggest that we fill this gap by drawing on Mills's (1997) conception of the Racial Contract. As Mills notes when discussing the Racial Contract, the point is not merely to provide a "factual/conceptual revision," but to provide an "alternate conceptual framework... a theoretical revision" that can be used as a tool for "making cognitively vivid the history of racial injustice, and then to facilitate... the discussion of matters of non-ideal rectificatory theory" (Pateman and Mills 2007, 237). Mills is explicit: "So for me, then, the contract has two dimensions, the descriptive/explanatory and the normative/prescriptive" (Pateman and Mills 2007, 231). To that end, I introduce the Racial Algorithm as the process by which histories of racial domination can be translated into data distributions, statistical models, and institutional decisions, producing new distributions that can then re-entrench the assumptions of the Racial Contract. By appealing to Mills, we can provide a structural account of why data-driven decisions can systematically overburden Black populations. This account not only diagnoses the mechanisms of injustice but also clarifies where interventions must occur if rectification is to be meaningful.

24.3.1 Defining the Racial Algorithm

Mills's concept of the Racial Contract provides the theoretical foundation for understanding how algorithmic systems can reproduce racial domination in contemporary society. Mills argues that the actual organizing agreement of modernity has been the Racial Contract rather than the idealized social contract described by liberal political theorists (Mills 1997). As Mills notes, the United States was never a "cooperative venture for mutual advantage" in Rawls's sense (Rawls 1999, 4). Mills demonstrates that modern Western societies were instead founded on Black enslavement and indigenous expropriation, with the purpose of "the differential privileging of the whites as a group with respect to the nonwhites as a group, the exploitation of [nonwhites'] bodies, land, and resources, and the denial of equal socioeconomic opportunities to [nonwhites]" (Mills 1997, 11). The Racial Contract shapes not only political institutions but also the norms by which legitimacy, morality, and knowledge are defined. It creates what Mills calls an "epistemology of ignorance"—a systematically produced misrepresentation of social reality that makes operations of power and exclusion invisible even as they continue to structure social relations (Mills 2017, 53–9). This epistemic distortion enables institutions to appear neutral while maintaining patterns of racial stratification.

The Racial Algorithm extends Mills's insight into the technical domain. Where Mills's The Racial Contract explains how social mechanisms organize white privilege and Black subordination, the Racial Algorithm is an explanatory model of how those mechanisms propagate through the machine learning pipeline. Causal mechanisms shaped by the Racial Contract generate data; supervised learning algorithms fitted to that data produce models whose statistical patterns are consistent with—but causally underdetermined by—whatever generated the training data; and those models inform decisions that produce new data, creating conditions under which the Racial Contract's effects can be compounded across iterations.

24.3.2 Data generation and the problem of "ground truth"

Machine learning systems depend on training data that mirrors existing social patterns. The "ground truth" observable distributions used to train algorithms are the product of causal mechanisms, including racialized ones, that generate them. In healthcare, for example, past medical expenditures may be

used as a proxy for health needs. However, expenditures reflect decades of segregated care, uneven insurance coverage, and discriminatory practices. This causal reality means that Black patients with identical conditions have historically received less costly care. Algorithms that optimize against such data can reproduce these inequities. The same dynamic appears across domains: Criminal justice algorithms trained on arrest and conviction data inherit patterns of racially selective enforcement; educational algorithms using standardized test scores reproduce the effects of segregated and unequally funded schools; employment algorithms trained on hiring and promotion histories perpetuate workplace exclusion.

A crucial choice point, then, lies in how to treat this so-called ground truth: As a neutral distribution to be replicated, or as a descriptively accurate yet morally compromised product of racialized institutions, thus requiring redress. To be clear, not all data distributions arise from unjust processes; in principle, they could reflect fair or cooperative causal mechanisms, and models themselves are agnostic as to whether patterns stem from just or unjust causes. However, in societies organized by the Racial Contract, data distributions that reflect racially differential adverse outcomes are more plausibly understood as products of domination than of an ideal moral order. In the absence of a compelling alternative causal story, the Racial Contract provides a plausible explanation for why these racial inequities persist. Accordingly, on a redress stance, “ground truth” distributions are objects for correction: Task choice, target definition, measurement (with invariance), optimization, and deployment must be disciplined by the obligation not to perpetuate injustice. Otherwise, “accuracy against the data” risks reproducing the residue of institutional practices shaped by colonial-era structures—including (but not limited to) enslavement, segregation, exclusion from wealth accumulation, differential policing, and unequal public investment.

24.3.3 Algorithmic training and the reproduction of hierarchy

With this “ground truth” distribution in mind, let us now turn to algorithmic training. In supervised learning, algorithms minimize a specified loss function relative to observed data patterns. Thus, when training data is the product of racial stratification, algorithms can produce models that effectively learn to predict outcomes reflecting those unjust patterns. The technical aspects of this process may obscure its normative significance. Loss functions, which measure the gap between predicted and observed outcomes, may seem like neutral technical choices but can in fact encode judgments about which types of errors matter and for whom. Many standard loss functions treat errors uniformly across groups, ignoring that different types of errors can carry very different consequences for different groups. Even incorporating statistical measures of fairness can still reinforce this problem by treating existing distributions as appropriate baselines. For example, calibration requires that predicted risk scores correspond to observed outcome rates consistently across groups—but this is consistent with the underlying observed outcome rates themselves reflecting just or unjust causal processes.

24.3.4 Institutional deployment and decision-making

Once trained, algorithms yield models that guide institutional decisions. Algorithmic decision-making does not operate in isolation. Algorithmic decision-making often operates within institutional contexts that shape how both private and public actors interpret and act on model outputs. To the extent that the Racial Contract shapes these institutional contexts, the resulting decisions can reflect not only statistical patterns but also the norms, conceptual resources, and epistemologies of racial domination. As such, even with technical bias mitigation techniques such as pre-processing data modifications, in-processing fairness constraints, or post-processing output calibrations, the Racial Algorithm can still produce normatively impermissible outcomes.

This persistence occurs through several mechanisms. First, the causal mechanisms underlying model outputs are underdetermined by the statistical patterns the model has learned; multiple distinct causal mechanisms can generate the same statistical patterns, but pattern recognition alone cannot identify which mechanisms are operating. Without understanding whether observed patterns reflect individual agency or structural constraints, decision-makers risk drawing normatively unwarranted inferences from statistically accurate predictions. Absent normative guidance, decision-makers may interpret this causal indeterminacy in ways that reflect, rather than resist, the Racial Contract. For example, judges may treat high recidivism risk scores as evidence of individual criminal propensity rather than as reflections of racially selective enforcement.

Second, unjust incentive structures can shape how institutions deploy algorithmic outputs. Some actors operate with objectives directly opposed to fairness—for example, deploying predictive policing to maximize asset forfeiture, or steering credit decisions toward high-fee products that burden low-income borrowers. Others may share fairness goals, but they operate within institutions whose norms, narratives, and accountability structures have been shaped by patterns of racial stratification. Thus, whether motivated by profit or fairness, decision-making risks remaining anchored in the Racial Contract.

24.3.5 Distributions, decisions, and compounding effects

As a result of these mechanisms, the Racial Algorithm can re-entrench initial disparities over time. Distributions generated by causal mechanisms under the Racial Contract are used to train algorithms, which produce models that inform decisions made by private or public actors. Some of these decisions take place outside of formal institutions and are shaped by broader cultural worldviews, conceptual resources, and norms. For example, judgments about who is risky, who is creditworthy, who is prepared for college, or who deserves opportunities. Other decisions occur within institutions that are themselves, in part, structured by the Racial Contract. In both contexts, the logics of racial domination can be reproduced, resulting in data-driven decisions that reflect and reaffirm racial hierarchies.

Once enacted, these decisions generate new data that re-enters the machine learning pipeline, potentially creating a lock-in effect that compounds the assumptions of the Racial Contract: Decisions made on the basis of algorithmic models can produce feedback that makes the underlying structures of racial domination increasingly durable and resistant to disruption. For example, in criminal justice, risk assessments shape sentencing and supervision, which in turn yield data portraying certain communities as higher risk, not necessarily because they pose greater threats but because they may face more opportunities for detection and sanctioning. Decision makers may then interpret this pattern as evidence of criminal propensity rather than as a reflection of selective enforcement, thus re-entrenching and reaffirming the worldview structured by the Racial Contract.

24.4 Ethics, anti-ethics, and algorithmic abolitionism

While incremental fairness reforms can sometimes reduce adverse outcomes at the margins, the persistence of racial disparities despite extensive technical fairness work raises a deeper evaluative question: How do we decide when continued reform is appropriate and when abolition is required? If an algorithmic decision-making system is deemed beyond repair, one might consider eliminating it. For example, on a global scale, we might consider banning the use of AI for lethal autonomous weapons in warfare (see the 2016 *Autonomous Weapons Open Letter* for an articulation of this stance). A domestic analog would be the abolition of what some characterize as coercive algorithmic systems, such as predictive policing, automated risk assessment tools in criminal justice, or other technologies

that can subject certain populations to heightened monitoring and control. Technical metrics alone cannot answer these questions, since algorithmic systems operate within institutions shaped in part by the Racial Contract. What is needed is a normative frame—one grounded in the structural insights of the Racial Contract—that can guide us in choosing between reform and abolition.

Mainstream debates about values in science have tended to rely on idealized conceptions of neutrality, procedural fairness, and democratic endorsement. However, a parallel body of work in Black philosophy and critical theory offers resources that have been largely overlooked but are potentially essential for grappling with the Racial Algorithm. In this section, I examine Leonard Harris’s insurrectionist ethics (2002), Tommy Curry’s anti-ethical critique (2011), Maggie O’Neill’s analysis of carceral logics (2025), and Daniel Fryer’s gradualist appeal (2020, 2023). Taken together with Mills’s non-ideal theory, these frameworks sharpen the tension between reform and abolition: While technical fairness-oriented reforms may mitigate immediate adverse outcomes, they may also reinforce systems whose structural function can be understood as domination.

24.4.1 *Insurrectionist ethics versus anti-ethics*

Harris (2002) grounds his insurrectionist proposal in the recognition that conventional moral philosophy fails to address the realities of racial domination. He insists that any moral system that does not affirm resistance as a duty is incomplete: “A philosophy that offers moral intuitions, reasoning strategies, motivations, and examples of just moral actions but falls short of requiring that we have a moral duty to support or engage in slave insurrections is defective” (2002, 192). This demand reframes ethics in terms of advocacy and resistance rather than abstract principles.

Drawing on David Walker’s (1829) Appeal, Harris emphasizes that the inclination to resist oppression is not optional but constitutive of human dignity: “the disposition to resist oppression and a correlative duty [to resist oppression] are facts of existential being” (Harris 2013, 96–7). Insurrectionist acts may rarely secure full liberation, but they are historically among the central means through which transformative change has occurred. As Harris argues, “there is no history of evolution without the history of insurrections, revolts, and revolutions” (2002, 203). The moral worth of insurrection lies not in its likelihood of success but in its assertion of dignity against systems whose function is designed to deny humanity.

Building on Harris, Curry’s anti-ethical framework advances a more radical challenge: On Curry’s account, it is a mistake to apply moral categories to institutions whose core function is domination rather than justice. Curry argues that traditional ethical frameworks commit a category error when they attempt to regulate racism through moral discourse, because racism itself operates outside the bounds of conventional morality (Curry 2011). As Curry explains, under colonial conditions, “the only ethical deliberation concerning racism must be anti-ethical, or a judgment refusing [to] write moralities onto immoral entities” (2011, 143, emphasis mine). This refusal is not moral nihilism but a recognition that institutions built on racial domination cannot be made moral through reform. The institution must be transformed or abolished.

Curry’s analysis reveals how moral discourse can function ideologically to legitimize systems of domination. When we apply ethical categories like fairness, accountability, or justice to historically racist institutions, we attribute to them “universal moral qualities that have yet to be demonstrated” (Curry 2011, 143). This abstraction from empirical reality allows us to maintain, “despite history or evidence, that racist entities will act justly given the choice” (Curry 2017, 183). As such, Curry argues that under conditions of sustained racial domination, ethical discourse itself becomes a form of idealist hope. The problem is not merely that conventional ethics fails to center the underrepresented groups, as insurrectionist ethics would have it, but that it creates false hope about moral transformation within dominant groups, which may paralyze resistance efforts (Curry 2017).

The relationship between Harris's insurrectionist ethics and Curry's anti-ethics reveals a crucial dialectical tension that illuminates the broader debate between reform and abolition in algorithmic governance. Both approaches emerge from recognition that traditional moral philosophy inadequately addresses structural domination, yet they differ in their orientation toward ethical discourse itself. Harris's insurrectionist ethics maintains that ethical frameworks may be reconstructed if properly grounded in the duty to resist oppression and centered on the dignity of oppressed peoples. In this sense, insurrectionist ethics represents what Urquidez calls "ante-ethics"—a condition of possibility for genuine moral community that must be established before ethical discourse, as conventionally practiced, can become (2023, 164). From this perspective, insurrectionist action creates the conditions under which oppressed individuals can have futures and thus participate in futural ethical deliberation. Curry's anti-ethics, on the other hand, denies racist institutions' moral status altogether.

This divergence creates what Urquidez identifies as a choice between "idealist" and "realist" ethical orientations. The idealist reaction maintains faith in moral suasion despite empirical evidence of its historical failure. In contrast, the realist reaction "refuses to engage in ethical deliberation with (and about) whites, thinking this pointless and unproductive given racism's apparent intractability" (Urquidez 2023, 166). The realist position does not abandon ethics entirely. Still, it adopts what Curry calls an "eschatological calculus that enables him to struggle against oppression, to live within tyranny, despite the seeming permanence of the repressive conditions" (2017, 186). The dialectical tension between these positions illuminates the space in which algorithmic governance debates occur. Technical reforms that promise to make systems more fair often operate within the idealist register, assuming that racist institutions can be morally transformed through appropriate procedural constraints. Anti-ethical analysis suggests that such reforms may serve to legitimize rather than challenge systems whose core function can be understood as racial domination.

24.4.2 Abolitionism versus gradualism

Maggie O'Neill's analysis of carceral logics provides an important insight into how algorithmic systems extend and transform historical patterns of racial control (2025). Carceral logics, as O'Neill describes them, are "the logics that organize institutions like the police, prisons, [and] the legal system, through a punish and control mindset" (2025, 19). These logics operate both informally, through ideologies and practices that naturalize punishment and surveillance as appropriate responses to social problems, and formally, through institutions such as police and courts. These carceral institutions, she argues, do not merely fail to deliver justice; they are built on a model of control incompatible with justice. For abolitionists, the goal is not reform but the creation of new democratic institutions that resist carceral premises; O'Neill characterizes abolitionism as "the approaches that problematize, challenge, and imagine beyond carceral logics and their manifestations across society" (22).

O'Neill argues that carceral logics extend beyond criminal justice institutions to shape broader patterns of social organization. They do so by providing "the motivations and assumptions" that frame social problems in terms of individual pathology to be punished rather than structural conditions to be addressed (2025, 21). In educational contexts, for example, carceral logics manifest through zero-tolerance policies, surveillance technologies, and disciplinary practices that criminalize student behavior rather than addressing underlying needs. As she puts it, carceral logics "inform not only practices in schools, but also how we are able to reflect upon, analyze, or imagine beyond these practices" (2025, 22, emphasis mine). This epistemic dimension may help explain why some technical approaches to algorithmic bias often fail to achieve meaningful change to the extent that they operate within conceptual frameworks that take punishment and surveillance as natural responses to social problems, making it difficult to imagine alternatives that prioritize care and community support.

In contrast, Fryer's work on criminal justice reform offers an alternative to both uncritical reform and absolutist abolition (2020, 2023). Fryer (2023) offers a sharp critique of abolitionism, accusing abolitionist discourse of relying on idealizations that can be counterproductive to addressing current policing problems. For example, the abolitionist ideal of a totally policeless society (563). He argues that the imaginative aspects of abolitionist rhetoric "risk being counterproductive with respect to abolitionists' activist goals" because they: "1) suggest that if abolition is desirable, approximating abolition is also desirable; 2) fail to clarify what the condition of applications of their view is; and 3) operate without the feasibility-sensitivity that one should expect from a non-evaluative position" (Fryer 2023, 569). On Fryer's view, abolitionist discourse lacks the kind of practical orientation necessary to build effective programs of change.

As an alternative, Fryer proposes a framework of repair. He envisions "a program that has desirable goals, clear applicability now, and sensitivity to feasibility constraints" (Fryer 2023, 570). A repair agenda, he notes, has often been dismissed by abolitionists as inadequate, but Fryer insists that "the goal here is to repair our society, not (just) the police." A reparative framework, he argues, would focus on "eliminating oppression, establishing institutions that allow us to be treated as equal under the law, and creating a society where persons relate to each other as equals in their day-to-day lives" (Fryer 2023, 570). By centering repair rather than speculative abolition, Fryer contends, we would be better positioned to achieve substantive changes without relying on "radical speculative imagination" (2023, 569, citing Roberts 2019). Similarly, in his 2020 paper, Fryer suggests that rather than treating reform and abolition as mutually exclusive strategies, we might understand abolition as a "gradual project of dismantling oppressive structures" (2020, 801). This gradualist approach recognizes that abolitionist goals cannot be achieved overnight and that reform efforts may be necessary to mitigate immediate adverse outcomes while advancing longer-term structural change. The key insight of Fryer's gradualist suggestion is that reform efforts must be evaluated not only by their immediate effects but also by their relationship to broader transformative goals.

Some critics of police abolitionism raise the "vulnerable populations" argument, which holds that eliminating certain institutions would leave already marginalized groups, especially Black people, more exposed to adverse outcomes. However, this line of argument can misrepresent the nature of policing as it actually operates in many practices. Police officers receive only a few months of training, far less than the three years of intensive legal education required for lawyers, who must also pass a bar exam before they are even allowed to argue cases in court. However, police, with this comparatively minimal preparation, are empowered to issue legal citations, to intervene in mental health crises often without background in psychology or counseling, to involve themselves in domestic disputes without training in conflict adjudication, and most significantly, to exercise lethal force at their own discretion. A police officer can decide in an instant to use lethal force. Unlike an ordinary citizen, that officer is not immediately subject to trial or judgment by a jury of peers. Ordinary citizens face immediate legal jeopardy for acts of violence; police, by contrast, are often shielded by doctrines of qualified immunity and by internal review processes that function as mechanisms of self-protection rather than democratic accountability.

The first line of accountability is typically internal, conducted by the department itself. Federal probes occur only rarely, and usually only when significant outside pressure compels them. Officers or their colleagues often determine whether charges will be filed in the first place. Even body cameras, only recently mandated, can be turned off by officers without meaningful consequences. From the abolitionist perspective, the police can function as a militaristic arm of the state, operating with significant impunity, empowered to take life and to determine for themselves whether that taking of life was "improper." To the extent that this characterization holds, the process is structurally undemocratic.

If the goal is public safety, then those functions can be separated and distributed across specialized institutions. Armed responses to active shooters could be handled by National Guard personnel or

military-trained specialists under strict civilian oversight. Trained conflict adjudicators could manage domestic disputes. Social workers and service providers could address mental health crises, homelessness, and substance use. Traffic enforcement could be automated through cameras, with citations mailed directly. More robust call centers could direct situations to the right professionals, ensuring that in most cases, unarmed responders—not armed officers—are dispatched. On this view, we do not need a single institution structured as contemporary policing is to address these diverse needs.

Critics like Fryer might respond that abolition is a risky gamble: It means replacing known, if flawed, institutions with alternatives whose efficacy is uncertain. The empirical evidence for proposed alternatives is limited, which might seem to counsel preserving existing arrangements. This gradualist caution frames abolition as premature, given the available evidence.

However, limited empirical evidence for alternatives is not a sufficient reason to preserve institutions that produce racially differential adverse outcomes. When evidence is incomplete, a plausible stance is pragmatic experimentalism: Evaluating existing institutions while testing alternatives. Even at some short-term cost, such experimentation may be necessary to prevent existing inequities from becoming self-reinforcing through the iterative pipeline described above.

Moreover, the prudential reasoning behind gradualism rests on a questionable assumption: Namely, that institutional change is a one-shot game. A one-shot game is a static model where actors move once, payoffs are realized immediately, and there are no further rounds. In such a framework, rational actors maximize short-term utility. Consider a bank deciding whether to extend loans to a group with poor credit histories. In a one-shot model, the bank's optimal strategy is to deny loans, since defaults threaten immediate payoff. Even if the bank values social uplift, lending remains irrational if it aims to maximize returns in a single round.

Borrower incentives compound the problem. Suppose group members receive a signal that the bank regards them as non-creditworthy and will likely deny loans. Since improving one's FICO score or building savings requires costly effort, it becomes rational not to invest if denial is expected regardless. The equilibrium here is negative: The bank rationally denies loans, the group rationally avoids investment, and creditworthiness does not improve. This equilibrium can also produce a double bind. On the one hand, denying credit harms those who might otherwise have been creditworthy, through no fault of their own, perhaps due to structural or historical factors. On the other hand, if the bank lent indiscriminately and collapsed under defaults, the group would also suffer, since the bank is itself an important community institution. In a one-shot framework, adverse outcomes can emerge either way.

The dynamic changes under a repeated game. In repeated play, outcomes in one round affect strategies in later ones. Here we assume—or allow—that the group's creditworthiness grows by some small amount α each round if they invest. If the bank is willing to extend loans in early rounds—even at a cost, then this changes the incentives. Once group members receive a positive signal that credit is genuinely accessible, they now have reason to invest in themselves despite the cost, because those investments will be rewarded in future rounds. As creditworthiness rises by α each period, default rates fall, and over time, the bank earns more than it would have by denying credit from the start. Crucially, this shift makes the equilibrium rational for both sides: It becomes rational for the bank to lend despite early losses, since long-run profits are higher, and for the group to invest, since the expected return on effort is no longer zero but compounded across rounds.

Notice that this change in rationality comes purely from the shift in time horizon. No external constraint was imposed on either agent's optimization problem. The bank is not “forced” to lend for fairness reasons, and the group is not “forced” to invest; both act rationally given their incentives.

The equilibrium changes simply because repeated play allows creditworthiness to grow over time. This differs from models that rely on constrained optimization, such as the affirmative action frameworks formalized in the Coate-Loury model (see [Coate and Loury 1993](#)), which depend on restricting

or reshaping choices. Here, rationality itself shifts without constraint once we abandon the one-shot assumption.

The same reasoning can extend to the question of abolition. The first rounds of transition away from policing may involve costs. Some individuals may be worse off temporarily than under the existing institution. However, if resources are reallocated into housing, conflict adjudication, mental health, and community infrastructure, then across repeated rounds, the capacity of communities can grow, just as creditworthiness grows in the bank example. Communities receive positive signals that investments in alternative structures will be rewarded, and this changes their incentives. As with the bank and the borrowers, the logic can become rational for both sides: Policymakers rationally invest in reallocations because long-term structural payoffs may exceed short-term stability, and communities rationally invest in new practices because their efforts are now recognized and reinforced. Short-term costs are not necessarily reckless gambles but can be understood as rational investments: They may be outweighed by the long-run structural benefits that policing, as currently structured, may not be able to deliver. Abolition, viewed through the lens of repeated play, can thus be understood as rational because it recognizes that long-term payoffs may depend on accepting and strategically managing short-term costs.

A further concern is that the modern digital environment may complicate the epistemic conditions under which gradualism is evaluated. Algorithms optimized for engagement can create epistemic bubbles and echo chambers that narrow collective judgment (Nguyen 2020), while widespread reliance on AI systems may produce what critics describe as moral and cognitive deskilling, eroding people's capacity for empathy, relational judgment, and democratic deliberation (Fang, Liu, and Danry 2025). To the extent that these dynamics obtain, they raise the question of whether we can reliably distinguish genuine institutional progress from the normalization of diminished baselines. This is not, in itself, a decisive argument against gradualism. Still, it suggests that any gradualist program must build in robust mechanisms for independent evaluation—mechanisms not themselves subject to the epistemic distortions it seeks to track.

In general, algorithmic systems exemplify the larger question of abolition or reform. As O'Neill argues, their capacity for continuous monitoring and predictive intervention can extend carceral logics, potentially sliding into anticipatory punishment that targets communities on the basis of statistical risk rather than behavior. From this vantage, technical reforms risk legitimizing domination by making punitive systems appear fairer. Fryer, however, urges a gradualist perspective: Reforms may sometimes be necessary to reduce immediate adverse outcomes, provided they are oriented toward dismantling rather than stabilizing oppressive institutions. The tension lies in determining when algorithmic reform functions as a step toward transformation and when it reinforces the very structures it purports to correct.

24.5 Toward a non-ideal theory of scientific legitimacy

The preceding analysis of the Racial Algorithm makes the central tension clear: Algorithmic reforms may both reduce immediate adverse outcomes and reinforce systematic racial inequities. Statistical parity, improved accuracy, or community consultation can make institutions appear more legitimate while leaving their core functions intact. However, refusing all reforms risks abandoning vulnerable populations to the adverse outcomes they face without mitigation. The central normative question is therefore not simply how to repair individual algorithms, but whether the institutions deploying them can be made more just. Framed this way, the reform-versus-abolition tension in algorithmic governance becomes part of a broader debate about democratic scientific legitimacy: Under conditions of structural injustice, the values that guide scientific inquiry—what aims to pursue, what targets to measure, what risks to tolerate—demand normative scrutiny.

Mainstream accounts of scientific legitimacy often rely on idealized democratic principles such as public reason, deliberation, or majoritarian endorsement (Badano 2025; Lusk 2021; Schroeder 2021; see Tolbert (forthcoming) for discussion). My critique is not that these theories fail to exclude values justifying racially differential treatment formally. Rather, it is that such formal constraints remain substantively perfunctory: They do not adequately theorize how historical and institutional asymmetries shape processes of value selection and claims to democratic legitimacy, even in the present. These approaches to legitimacy repeat a foundational mistake that Mills diagnosed within liberal political theory: They omit “crucial facts” about the histories of racial domination that structure our epistemic and political institutions (1994, 226). Anita Allen makes a similar point in her discussion of the Black Opticon, her name for the overlapping systems of over-surveillance, exclusion, and exploitation directed at the Black community: “generic calls on behalf of all population groups are insufficient to shield the African American community from the Black Opticon,” making the shift from general reform to “race-conscious reform” essential (2022, 912). The implication is clear: Legitimacy cannot be theorized solely in terms of abstract ideals; it must be grounded in the structural asymmetries that condition both data and decision-making.

24.5.1 Criteria for a non-ideal theory of algorithmic governance

The non-ideal framework of scientific legitimacy I defend here has two components. First is a procedural: Every publicly funded project that sorts, screens, or surveils people must (1) publish a prospective disparate-impact audit before deployment, (2) release a retrospective audit after deployment, (3) maintain a continuously updated public docket of data sources and model choices, and (4) recognize a community-initiated veto that pauses funding or rollout whenever those poised to bear the greatest risk file a petition.

During the pause, investigators shoulder a burden-shifting obligation: They must demonstrate (1) a compelling public or scientific benefit, (2) the absence of any less-harmful alternative, and (3) that the project will not exacerbate existing inequalities or produce disproportionate adverse outcomes for certain groups. An independent panel—half appointed by the funding agency, half drawn from a roster nominated by community boards—issues a public ruling to lift, modify, or extend the suspension.

Second, a substantive component specifies which kinds of values may legitimately steer scientific work. Derived from Mills’s corrective justice paradigm, the substantive core contains four interlocking norms: (1) Recognition of the equal moral status of all people, (2) prohibition of domination and exploitation, (3) a reparative orientation toward unrepaired historical harms, and (4) epistemic pluralism that refuses to dismiss community knowledge on purely arbitrary grounds. The equal moral status norm rules out any method that prices one group’s adverse outcomes more cheaply than another’s; the audits make that valuation explicit by forcing researchers to publish group-specific error costs and keep them within the same range. The anti-domination/anti-exploitation norm blocks models that would worsen the position of already disadvantaged groups, even if the models’ deployment could be justified by reference to overall social benefit. The reparative orientation norm flips the evidentiary burden: If a dataset reflects the legacy of racial discrimination or structural racism, whether through redlining, segregated medical spending, or racially selective arrests, among other mechanisms. This goal is to have investigators justify their continued reliance on that legacy distribution and explain how their project will shrink, rather than reinforce, the gap.

Both components are important. Procedural safeguards without substantive constraints risk reducing legitimacy to box-checking, while substantive norms without procedural mechanisms risk technocratic imposition. Only their combination ensures that algorithmic governance is both democratically accountable and normatively constrained. Together, these criteria can increase the likelihood that reform efforts are substantively just. If an institution persistently cannot meet even these

minimal procedural and substantive demands, then technical adjustments are more plausibly understood as legitimation than as justice. The framework thus provides a principled basis for evaluating cases where algorithmic reform can reduce adverse outcomes and for identifying conditions under which the institutional structure itself may call for more fundamental restructuring.

24.6 Conclusion: Abolitionist horizons

I propose that this chapter contributes to values-in-science debates in three ways. Conceptually, it shifts focus from isolated algorithmic fairness metrics to a normative explanatory model of racial stratification. Normatively, it supplies criteria for the non-epistemic values that ground scientific inquiry. Institutionally, it specifies non-ideal procedural and substantive requirements for legitimate algorithmic governance, thereby offering guidelines for evaluating whether reform or more fundamental institutional transformation is appropriate. The upshot is that algorithmic governance cannot be assessed solely by reference to technical performance or abstract fairness metrics, and values in science cannot be theorized apart from histories of domination. Fairness metrics and incrementalist fixes address symptoms, but scientific legitimacy requires confronting causes: The racialized institutions, data, and logics that shape scientific practice.

Note

- 1 I credit Aaron Roth with this example in a Slack conversation.

References

- Allen, Anita L. 2022. "Dismantling the 'Black Opticon': Privacy, Race Equity, and Online Data-Protection Reform." *Yale Law Journal Forum* 131: 907–58. https://scholarship.law.upenn.edu/faculty_scholarship/2803.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias." ProPublica, May 23. www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
- Autonomous Weapons Open Letter: AI & Robotics Researchers. 2016. Open letter, February 9. <https://futureof-life.org/open-letter/open-letter-autonomous-weapons-ai-robotics/>.
- Badano, Gabriele. 2025. "Public Reason, Values in Science, and the Shifting Boundaries of the Political Forum." *Philosophical Studies* 182: 1133–55. <https://doi.org/10.1007/s11098-025-02323-x>.
- Chouldechova, Alexandra, and Aaron Roth. 2020. "A Snapshot of the Frontiers of Fairness in Machine Learning." *Communications of the ACM* 63 (5): 82–9. <https://doi.org/10.1145/3376898>.
- Coate, Stephen, and Glenn C. Loury. 1993. "Will Affirmative-Action Policies Eliminate Negative Stereotypes?" *American Economic Review* 83 (5): 1220–40. <https://www.jstor.org/stable/2117558>.
- Curry, Tommy J. 2011. "The Political Economy of Reparations: An Anti-Ethical Consideration of Atonement and Racial Reconciliation under Colonial Moralism." *Race, Gender & Class* 18 (1/2): 125–46. <http://www.jstor.org/stable/23884871>.
- Curry, Tommy J. 2017. *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood*. Temple University Press.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Fang, Cathy Mengying, Auren R. Liu, Valdemar Danry, et al. 2025. "How AI and Human Behaviors Shape Psychosocial Effects of Chatbot Use: A Longitudinal Randomized Controlled Study." *MIT Media Lab*, March 21. Accessed September 10, 2025. <https://www.media.mit.edu/publications/how-ai-and-human-behaviors-shape-psychosocial-effects-of-chatbot-use-a-longitudinal-controlled-study/>.
- Fryer, Daniel. 2020. "Race, Reform, and Progressive Prosecution." *The Journal of Criminal Law and Criminology* 110 (4): 769–802.
- Fryer, Daniel. 2023. "Idealizing Abolition." *Criminal Law and Philosophy* 17: 553–72. <https://doi.org/10.1007/s11572-023-09700-9>.
- Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *The Journal of Finance* 77 (1): 5–47. <https://doi.org/10.1111/jofi.13090>.

- Harris, Leonard. 2002. "Insurrectionist Ethics: Advocacy, Moral Psychology, and Pragmatism." In *Ethical Issues for a New Millennium*, edited by John Howie, Southern Illinois University Press.
- Harris, Leonard. 2013. "Walker: Naturalism and Liberation." *Transactions of the Charles S. Peirce Society* 49 (1): 93–111. <https://doi.org/10.2979/trancharpeirsoc.49.1.93>.
- Hellman, Deborah. 2025. "Algorithmic Fairness." In *The Stanford Encyclopedia of Philosophy*, Fall 2025 Edition, edited by Edward N. Zalta and Uri Nodelman.
- Kleinberg, Jon, Sendhil Mullainathan, and Manish Raghavan. 2016. "Inherent Trade-Offs in the Fair Determination of Risk Scores." <https://doi.org/10.48550/arXiv.1609.05807>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- McPherson, Lionel K. 2024. *The Afterlife of Race: An Informed Philosophical Search*. Oxford University Press.
- Mills, Charles W. 1994. "Non-Cartesian Sums: Philosophy and the African-American Experience." *Teaching Philosophy* 17 (3): 223–44.
- Mills, Charles W. 1997. *The Racial Contract*. Cornell University Press.
- Mills, Charles W. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana. State University of New York Press.
- Mills, Charles W. 2017. *Black Rights/White Wrongs: The Critique of Racial Liberalism*. Oxford University Press.
- Mills, Charles W. 2020. *Theorizing Racial Justice*. Transcript of Tanner Lecture. University of Michigan, February 12.
- Nguyen, C. Thi. 2020. "Echo Chambers and Epistemic Bubbles." *Episteme* 17 (2): 141–61.
- O'Neill, Maggie Beth. 2025. "Educational Work Beyond Carceral Logics—Stereoscopic Engagements with Educational Philosophy and Abolitionism." PhD diss., University of Gothenburg. <https://hdl.handle.net/2077/86560>.
- Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–53.
- Pateman, Carole, and Charles W. Mills. 2007. *The Contract and Domination*. Polity Press.
- Rawls, John. 1999. *A Theory of Justice*. Revised ed. Harvard University Press.
- Roberts, Dorothy E. 2019. "Abolition Constitutionalism." *Harvard Law Review* 133: 1–122.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Tolbert, Alexander Williams. forthcoming. "The Racial Algorithm: Reassessing Scientific Legitimacy in a Value-Laden World." *Critical Philosophy of Race*.
- Urquidez, Alberto G. 2023. "Anti-Ethics as Insurrectionist Ethics: An Analysis of the Normative Foundations of Philosophies Born of Struggle." In *Insurrectionist Ethics: Radical Perspectives on Social Justice*, edited by Jacoby Adeshei Carter and Darryl Scriven, 157–94. Springer Verlag. https://doi.org/10.1007/978-3-031-16741-6_8.
- Walker, David. 1829/1995. *Walker's Appeal, in Four Articles; Together with a Preamble, to the Coloured Citizens of the World, but in Particular, and Very Expressly, to Those of the United States of America*. Hill and Wang.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

PART IV

Values, Science, Institutions, and Organizations



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

THE VALUES OF SCIENCE FUNDING INSTITUTIONS

Jamie Shaw

25.1 Introduction

Science funding policies are diverse and complex, partially owing to the multitude of values guiding or constraining them. These values can be genuine or public relations antics, widely accepted or controversial, or complementary or in tension. Here, I provide a rough-and-ready picture of the values of various science funding institutions while analyzing how they emerge, interact, and can be interpreted. This will be accomplished by analyzing the behaviors of science funding agencies while consulting existing literature on the subject, as well as commentaries about what values they should embody or prioritize. On occasion, this also involves adding my two cents. This, I hope, offers preliminary steps toward a greater understanding of the value-laden dimensions of science funding institutions. This analysis complements and extends philosophical discussions on science and values in two ways. First, while traditional discussions of science and values have focused on individual scientists, recent calls advocate increasing attention to institutional values ([Douglas 2018](#)). This analysis heeds this call. Second, some have been interested in questions about values in pursuit ([Elliott and McKaughan 2009](#); [Okruhlik 1994](#)). By looking at institutional values associated with funding, we can consider systematic issues impacting scientific pursuits.

In this chapter, I focus on some commonly held values of public, commercial, and philanthropic science funding institutions. These analyses reflect recurring states of affairs, rather than universal features of science funding policy. Consequently, readers should not assume that specific cases will necessarily reflect these analyses.

25.2 Public science funding

Some values concern what research institutions should support or prioritize (see, e.g., [Solovey 2020](#)). The corresponding “missions” might be broad (e.g., the European Research Council supports “investigator-led research”) or narrow (e.g., the African Centre for Aquatic Research and Education supports research on freshwater ecosystems within Africa). While the ultimate goal of public science funding institutions is, arguably, supporting successful science, this is not always their chief proximate goal. For example, Chandra [Mukerji \(1990\)](#) shows how American federal funding of oceanography in the late twentieth century primarily aimed to cultivate a class of scientists who could be used as spokespeople for future policies. Whether the science was valuable was incidental; what

was more important, Mukerji argues, were power relations and the perceived authority of particular public policies. Other common proximate goals include training, diversification, economic stimulus, or promoting collaborations. Debates have also arisen about what constitutes “success,” questioning whether broader societal impacts define “success” or “failure” or if scientific criteria are sufficient (Elliott 2017; Holbrook 2005, 157). Differing proximate goals make some values more salient while others fade into the background.

My focus here will be on values embedded in institutional day-to-day operations. Synthesizing sparse literature on this topic, there are nine values that public science funding agencies typically aim to embody:

- 1 Accountability: There should be systems of oversight that allow them to be rewarded, punished, or transformed.
- 2 Autonomy: They should have some independence from governments in policymaking and allow some independence for prospective applicants.
- 3 Trust: They should establish fiduciary relationships with the relevant public.
- 4 Transparency: Their norms, rationales, and mechanisms for engagement should be easily accessible.
- 5 Fairness: Funding evaluation mechanisms should minimize sources of discrimination.
- 6 Efficiency: Achieving institutional aims should incur minimal costs.
- 7 Reliability: Grant evaluation procedures should dependably identify proposals that best serve institutional interests.
- 8 Consistency: Application requirements should avoid drastic changes.
- 9 Diversity, Equity, and Inclusion (DEI): They should promote DEI within science.

25.2.1 Accountability

In democratic countries, public science funding agencies are accountable to elected representatives (Baldwin 2018; Douglas 2024), although in some cases, it can mean direct accountability to stakeholders (Elliott 2017, 29). Accountability usually involves back-and-forths between government and agencies (see below), so it is usually negotiated, even in non-democratic countries (e.g., Iran, see Heshmati and Dibaji 2019). What is negotiable usually depends on legal statutes. Accountability can involve budget reports, mission statements, job descriptions for high-ranking administrators, disclosure policies, or rules for external consultants. Some statutes allow for more invasive decisions. For example, the National Institutes of Health (NIH) has its review criteria dictated by Congress, although NIH administrators decide on their interpretation (Kozlov 2022). What accountability requires can be controversial, and in tension with other values—especially autonomy (see Elliott 2017; Weingart 1999, 27ff). The devil is usually in the details for how to balance these values (Cozzens 1999).

25.2.2 Autonomy

Autonomy is often perceived to be necessary for a few reasons. First, politicians usually have vested interests in siphoning funds into their jurisdictions, causing pork-barreling, even if it hinders scientific progress (Martino 2017). A related, more controversial consideration is whether politicians have relevant expertise. Consider, for example, Tom Coburn’s (R-Oklahoma) 2011 “NSF: Under the Microscope” report. Here, opportunistic and superficial accusations of wastefulness at the National Science Foundation (NSF) encouraged greater oversight (see also Pamuk 2019, 1). This shows, for some, that the ignorance or propaganda of politicians shouldn’t influence funding priorities (Elliott 2017, 31). Rather, science funding institutions usually want differently situated people (e.g., scientists) at the helm. Similarly, agency administrators usually want to protect contentious research (e.g., stem cells,

genetic screening, gun control) that might be banned or truncated by oversight committees. Some see restrictions on contentious research as constitutive of socially responsible science (Feyerabend 1991, 500). Others see this as an infringement on scientists' ability to speak independently (Wilholt 2010). Regardless, systems of accountability are more complicated than momentary decisions to avoid worrisome research. Legal statutes must consider diverse values that might steer science in different ways. Protecting institutions from liberal politics protects them from conservative politics and vice versa.

Oversight committees often want to curb autonomy to increase inter-agency coordination. This results in tugs-of-war where agency autonomy is pitted against broader agendas (Lepori 2011). Researcher autonomy *from* science funding agencies is frequently valued, as fully "top-down" approaches to research priorities are unrealistic and unwise (Aagaard et al. 2021). Allowing "bottom-up" approaches driven by researcher interests is often necessary. Balancing top-down measures (e.g., the Obama administration's "cancer moonshot") and bottom-up mechanisms (e.g., "fund people, not projects") is usually targeted.

25.2.3 Trust

Trusting science funding bodies involves believing that they will use their autonomy to serve the public's best interests. Science funding agencies rely on public support to maintain a healthy budget or autonomy, incentivizing them to hype successful or promising research they funded. Disillusioned and loud public can pressure reforms of institutional norms or research priorities. For example, Senator William Proxmire ridiculed projects funded by the NSF with publicly popular "Golden Fleece Awards" for their perceived uselessness, which almost led to the Baumann amendment that would have provided Congress with additional oversight powers (Baldwin 2018; Solovey 2020, Chapter 5). Trust is usually framed in generic terms concerning, e.g., the "value of science," since they usually struggle to detail direct social impacts from their funded research (Podhora, Helming, and Adenäuer 2013). Some agencies can be more vulnerable, such as NASA, given contentious public perceptions of space science (Steinberg 2011). Sometimes, when agencies fund socially controversial research, it can affect trust in the institution as a whole, such as dual-use research with military applications funded by Japan's Acquisition, Technology and Logistics Agency (Mallapaty 2022). Finally, people with historically tumultuous relationships with scientific institutions are often valuable or necessary partners in research or policy formulation. Repairing and promoting trust is often necessary for these initiatives to succeed (see, e.g., Manning and Walton (2021) for a discussion of Māori populations and science funding in New Zealand).

25.2.4 Transparency

Transparency is essential for the sheer existence of science funding institutions in countries struggling with corruption, like Poland currently (Brent 2024). Transparency often involves clarifying, for applicants, institutional norms or policies (e.g., agency priorities, evaluation criteria, and points of contact for questions or complaints). Transparency, for some, requires making successful and rejected grant applications and reviewer reports public (Horbach, Tjldink, and Bouter 2022). Perceived lack of transparency can foster resentment and frustration (see, e.g., McGinn et al. 2019), prompt scientists to leave the profession (Etzkowitz and Ranga 2011), refrain from applying for funds (Kanji 2015), or seek support from other agencies promoting different research (Cunningham et al. 2016; Myers 2020; cf. Diamond 1999). Transparency is often necessary for legitimate accountability mechanisms, and inadequate transparency can lead to decreases in autonomy or trustworthiness. Sometimes, science funding agencies disguise policies they deem necessary but may decrease public trust, thereby compromising on transparency.

25.2.5 Fairness

Concerns about fairness have become especially pronounced over the past 30 years amongst science funding policy commentators. Several studies show that peer review, the primary mechanism for grant evaluations, often unjustly favors researchers from particular social groups (Guthrie, Ghiga, and Wooding 2017). This perpetuates funding gaps, causing various down-the-road effects on science (see, e.g., Hall, Vogel, and Huang 2018). What counts as “fair” is complex. It requires distinguishing relevant from irrelevant applicant features (e.g., race, gender), where unfairness involves systematic deviances concerning the latter. However, in some contexts, applicant features are relevant, such as affirmative action funding policies or when a researcher’s background uniquely positions them for conducting research (Stevens, Masters, and Imoukhuede 2021). Therefore, we shouldn’t rush to a substantive notion of fairness independent of particular funding schemes. Fairness often imposes responsibilities on funding agencies, such as keeping detailed applicant, reviewer, and funding-recipient demographic data or experimenting with potentially fairer funding mechanisms (Sato et al. 2021).

25.2.6 Efficiency

Beyond operating costs, the growing concern is the opportunity costs of peer review. One study of the National Health and Medical Research Council of Australia calculated that 550 working years went into preparing grants for a single year of applications (Herbert et al. 2013). These opportunity costs are growing exponentially. This gives scientists less time for tasks, including reflection, public engagement, or mentoring. The pressures of peer review even interfere with scientists’ personal lives, affecting their mental health and relationships with friends and family (Isaacs 2016). These costs are usually invisible to funding agencies, since they are not borne by their personnel, but there is increasing external pressure to manage these imposed costs. Moreover, unevenly distributed costs lead to uneven applicant rates (von Hippel and von Hippel 2015). Specifically, early-career researchers, women, some underrepresented minorities, interdisciplinary researchers, and researchers at “less prestigious” universities are more frequently discouraged from applying for grants. This promotes a vicious cycle where fewer applications promote marginalization, promoting unfair funding distributions.

25.2.7 Reliability

Reliability is often a primary focus of science funding agencies, although measuring the reliability of funding mechanisms is controversial (Lindner and Nakamura 2015). Simply put, science funding agencies want mechanisms that dependably identify research that best advances their missions. Reliability comes in different guises, reflecting more specific values. Reliability of *individual* grant evaluations is more salient for institutions like the Office of Science at the Department of Energy that only fund a few projects. More commonly, “reliability” is construed in social terms where the primary targets of evaluation are research portfolios as wholes (Aagaard et al. 2020). Reliability can also be defined as minimizing false positives (wasting funds) or false negatives (missed opportunities). This can tie into trustworthiness, as perceived wastefulness pressures funding agencies to minimize false positives. Conversely, for military agencies, false negatives are far more significant, making them less risk-averse (Collins 2000, 832–33). The relationships between efficiency, fairness, and reliability in peer review are still being investigated. Lotteries, it is often claimed, increase the first two values and reliability qua minimizing false negatives, although they may increase false positives (Fang and Casadevall 2016).

25.2.8 Consistency

Scientists rely on steadily available funds (O’Grady 2022). Continuously available funds and consistent parameters of funding schemes, which require significant effort to master, are crucial for

maintaining applicant rates and keeping scientific projects alive. Consistency also involves implicit value judgments, particularly a conservatism that avoids dramatic experimentation with alternatives to the status quo. This makes science funding agencies, like most bureaucracies, slow to change, thereby decelerating progress on fairness, efficiency, and/or reliability. Consistency can also, in roundabout ways, be connected to fairness; extreme budget cuts in Brazil, for example, exacerbated gender inequality (Hipólito et al. 2022). For some, this value is sufficiently upheld by the constancy of institutional norms and available funding. For others, consistency entails more radical shifts toward non-competitive funding distributions to ensure a sufficient baseline for researchers to pursue science across longer time-horizons (Dresler 2023; Myers and Tham 2023).

25.2.9 Diversity, equity, and inclusion

Despite setbacks under the second Trump administration, it has become increasingly acknowledged that DEI initiatives are constitutive of fairness and necessary for scientific progress (Hofstra et al. 2020; Juk, Salles-Filho, and Pinto 2023). DEI can be promoted through incentives to include DEI in proposals, eligibility constraints on grant calls, or diversifying review panels (Kozlov 2022, 602). We know little about how DEI initiatives will impact trust in different social-political climates or how diversifying review panels impacts their reliability or fairness, but some are optimistic that these values can be jointly maximized (Taffe and Gilpin 2021).

25.3 Science for sale

Some see the increased presence of commercial science funding sources as a threat to socially responsible science (Fernández Pinto 2020). Others see commercial entities as having more appropriately situated infrastructure to ensure that science can be genuinely useful (Diamond 2006; Hong et al. 2015; Kealey and Al-Ubaydli 2000; Nishtar 2004). Though the line between public and commercial funding can be blurry (Chalmers and Nicol 2004), I focus here on some values in commercial science funding. Before proceeding, a few general observations are needed.

Commercial companies can be Janus-faced by claiming to adhere to some values while embodying others. Although this phenomenon exists in public or philanthropic settings (Edwards and Roy 2017; Krimsky 1999), it is especially pronounced in commercial contexts. For example, Pfizer’s “Transparency in Clinical Trials,”¹ emblazoned on their website, claims a commitment to transparency, including publishing negative results. But Pfizer, like most pharmaceutical companies, does not disclose all their collected data or publicly register all of their performed trials even when legally required (Sismondo 2018, 46). While many companies claim to be transparent, they usually fight tooth and nail to prevent internal documents like memos from being publicly released in court (cf. Plottel et al. 2020). Commercial science funding values may seem less pluralistic than other agencies, with profit as their ultimate value guiding other values. However, only sometimes do commercial institutions act like cartoon villains who will do anything for profit. Corporate ethea, although inconsistently adhered to, influence how research gets pursued (see Maienschein et al. 2008).

This being said, there are especially pronounced values in commercial settings. Here, I discuss the following:

- 1 Cost-Effectiveness: Research costs should be as minimal as possible.
- 2 Timeliness: Research should yield results on timelines that support company growth.
- 3 Integrity: Research should appear to be consistent with the values of independent science.
- 4 Accountability: Companies should appear to provide needed or desired goods and services.
- 5 Compliance and Trust: Publics should be willing to support or accept public company actions.

The first two values I present as genuine. Points (3) and (4) are more complex, as I will discuss the divergence between their marketing and implementation. The final value lies somewhere in between, depending on the situation.

25.3.1 Cost-effectiveness

The profit-motive prioritizes cost-effectiveness in commercial settings. This encourages investments tailored toward known commodities (i.e., lowering false positives in the sense used above) while letting public scientists or smaller firms take more epistemic risks (see [Aronson and Green 2020](#)), although this is not always the case for larger companies. While the exact dimensions of how pursuit worthiness judgments will be impacted by AI are an open question, commercial companies often penny-pinch more than public funding agencies.

25.3.2 Timeliness

Unlike public science, where progress is rarely directly held accountable (interim reports for public funding agencies are often a façade), sales quotas must be met regularly. This is especially the case for publicly traded corporations, whose shareholder meetings often require quarterly or biannual profit statements. This rushes science to meet deadlines, increases the intensity of promising scientific research programs, or prioritizes more tractable research.

25.3.3 Integrity

Scientific integrity is a frequently touted value in commercial settings since, unlike most publicly funded science, commercial funders are *actively involved in the science* (e.g., writing papers, preparing presentation slides) ([Sismondo 2021](#), 2–3). Scientists recruited by private companies often believe that they adhere to high standards of scientific integrity ([Sismondo and Chloubova 2016](#)). These beliefs are likely sincere, but rarely accurate, as they are often unaware of circumscriptions of their autonomy. To be clear, research is rarely completely fraudulent. Rather, researchers' degrees of freedom—conventional choices that are permitted by standard scientific norms—are chosen to facilitate company goals. [Elliott \(2014\)](#) argues that worries about commercial scientific integrity tend to be more significant under specific conditions (e.g., higher uncertainty, significant financial stakes) and less significant when these conditions are absent. Because of this, commercial companies can often legitimately say that they are acting with integrity. However, since their choices are often unaligned with broader values, integrity might not be truly valuable, all things considered.

25.3.4 Accountability

The place of commercial science in democratic societies is a complicated issue ([Bravo 2026](#)). Some see commercial companies as threats to democracy, as their outsized power allows them to propel self-interested policies at the expense of broader public engagement. General suspicions of corporate intentions encourage public relations where they claim to be acting toward the social good ([Benson-Greenwald et al. 2023](#)). Broadly speaking, commercial institutions often claim to be accountable to democratic norms. However, accountability is not construed as political oversight—although companies frequently advertise their compliance and cooperation with legal administrations—but in terms of *service*. In other words, “accountability” is about providing products that people need or want (cf. [Pulkkinen 2026](#) on their notion of “output legitimacy”). While public scientists sometimes use this framing when discussing social impacts, it is more prominent in commercial science with tighter connections between science and application.

25.3.5 Compliance and trust

Oftentimes, companies pursue compliance over trust. By “compliance,” I mean a willingness to be complicit in their actions. This might involve buying, supplying, or promoting their products, being quiet in public discourse about their impacts on research, or allowing them to shape regulatory policies. Compliance often falls short of trust but is often sufficiently valuable for commercial entities to pursue. Lack of compliance or distrust can lead to obstacles for commercial interests. For example, public perceptions of private industry impacted the negotiations of private to public partnerships when developing COVID vaccines (Akomea-Frimpong et al. 2023; see also Herder 2008). Compliance or trust can also impact sales in cases where consumers have genuine alternatives. To achieve compliance or trust, companies might perpetuate disinformation (Oreskes and Conway 2010), which could foster doubt that prevents active resistance or enable trust, or disproportionately flood information markets with preferred claims to the same effect (Legg, Hatchard, and Gilmore 2021). In some cases, compliance from the perspective of the company must be perceived as trust from the perspective of the public. For example, participation in human trials or focus groups often requires perceived overlap in values, or at least a lack of conflicts of interest, to flourish (Critchley, Nicol, and Otlowski 2015). Genuine trust has its own sales pitch: It secures longer-lasting relationships that are more reliable than those that might get damaged through information leakage or public criticisms.

25.4 Philanthropic science

At the turn of the twentieth century, philanthropic institutions were the largest funders of science in many parts of the world, especially in the US and parts of Northwestern Europe (Kevles 1992). While this philanthropic funding was mostly for “basic research,” this function has been largely supplanted by public funding agencies. As a result, most philanthropic science has shifted to ameliorating concrete problems (see Bravo and Elliott 2025). Like commercial funding institutions, their primary targets are usually interventions rather than “knowledge.” But unlike commercial funding, their primary value is not financial, prompting different values. The ones I discuss are as follows:

- 1 Stakeholder Involvement: They should define, seek out, and engage with stakeholders.
- 2 Sustainability: For long-term cases, they should stimulate goods or services available for the long term.
- 3 Political Legitimacy: They should be cooperative with legitimate governmental and/or intergovernmental institutions.
- 4 Trust and Compliance: They should secure the willingness to participate from patrons and stakeholders.

25.4.1 Stakeholder involvement

Philanthropic science funding agencies, being closer to the action than publicly funded science, typically need to identify and connect with stakeholders. Stakeholder engagement within public science is rare, but growing (Bloch, Ramos-Vielba, and Thomas 2023), and is often a mirage in commercial settings (Sismondo 2018, 163–7). Moreover, commercial companies often define stakeholder groups broadly to maximize their consumer base, rather than identifying genuine needs.

Philanthropic approaches to stakeholder involvement are mixed (Arco-Castro et al. 2020). Some “top-down” approaches determine stakeholders through internally generating reasoning, while others are more collaborative (Mati 2017; see also Elliott 2017, 139–48). Choosing stakeholders involves judgments about *who* matters more (Backer, Smith, and Barbell 2005) and what the *problem* is, which is closely connected to how the research should be performed (Michelson 2021). This final

point often blurs distinctions between science, politics, and other walks of life. Unlike public agency mission statements, which usually use disciplinary terms (e.g., “science,” “medicine”) to define their jurisdictions, philanthropic institutions typically have broader mission statements. To illustrate, the Paul K. Feyerabend Foundation’s (PKFF) statement is:

The [PKFF] promotes the empowerment and wellbeing of disadvantaged human communities. By strengthening intra and inter-community solidarity the Foundation strives to improve local capacities, promote the respect of human rights, and sustain cultural and biological diversity.²

Despite the fact that many projects funded by the PKFF are recognizably “scientific,” the term and its cognates are unused. Rather, targeted interventions—e.g., improving local capacities, promoting human rights—do the boundary work. Unlike public science institutions, where links between science and application are typically long, complex, or un(der)considered (Feller 2022), the values of philanthropic science are usually directly connected to targeted interventions. As has been well-documented, actionable science requires different methodological and theoretical choices, leading to different micro-values informing the nitty-gritty of scientific practice (Douglas 2000; Elliott 2011; Epstein 1996; Toomey 2024). Given the complexity of actionable science, dialogue is prioritized over policy, which pushes toward stakeholder engagements.

25.4.2 Sustainability and philanthrocapitalism

An increasingly prominent approach in modern philanthropic science funding is “philanthrocapitalism” (Jain 2012). The primary goal of philanthrocapitalism is to create, maintain, and/or manage *markets* by encouraging commercial or government investments into collaborative efforts to produce cost-accessible goods or services for target populations. I find it difficult to discern whether philanthrocapitalism requires commitments to “capitalist values” (e.g., profit, merit, competition) (cf. Barnett 2023). Critics often assume this about philanthrocapitalism, but philanthrocapitalism might be based on beliefs about efficiency. This obscured dialogue has prevented closer inspection of more specific value judgments in philanthrocapitalist methods. Regardless, the longer-term goal of creating markets fosters valuing sustainability significantly. Debates about the value of philanthrocapitalism are currently raging (Adewoye et al. 2023; Harvey, Gordon, and Maclean 2021). Critics worry that the profit-motive, beyond attracting bad actors, incentivizes the wrong kinds of interventions. In public health initiatives, for example, philanthrocapitalism is accused of unduly promoting pharmaceutical over social interventions. Defenders argue that philanthrocapitalism stimulates investment into areas that are systematically underfunded. Many environmental and medical problems plaguing large populations in the Global South receive little attention from commercial (Reiss and Kitcher 2009) or publicly funded scientists (Evans, Shim, and Ioannidis 2014). Publicly funded research on these problems often falters, given various entrenched political and institutional barriers for performing the research adequately (Aerts et al. 2017; see also Elliott 2017, 34–5). Therefore, investments must be stimulated in ways beyond most public or purely commercial approaches.

25.4.3 Accountability

Larger philanthropic bodies occupy confusing places in systems of political legitimacy (McGoey 2021). For example, the Gates Foundation has significant power to negotiate what systems of accountability they should be privy to (they are the second largest funder of the WHO). This is especially complex in

international cases, as philanthropic institutions can often act in ways that public agencies, due to issues of political sovereignty, cannot. This becomes further complicated when local governments are repressive or corrupt, making it questionable to whom they should be accountable. Still, it is recognized that oversight of philanthropic activities is necessary to ensure their political legitimacy (Weinryb 2020).

25.4.4 Trust and compliance

Philanthropic science funders usually require trust and compliance from different groups. Trust that their actions will accord with the values of potential donors, for example, is often necessary for obtaining funding (Chaudhry, Dotson, and Heiss 2021). Regarding those groups, philanthropic science funders aim to serve, achieving warranted trust is often difficult to secure in the short-term, since they have insufficient time to fully acclimate themselves, local populations can rarely seriously consider alternatives, and deeply entrenched differences in beliefs about the world are difficult to overcome quickly (e.g., transmission of STIs). Because of this, what we often see is an emphasis on compliance and “education.” This involves balances of values—expediency (via compliance) over trust. This, of course, is not always the case—especially in situations involving longer-term partnerships—but can be a practical necessity.

25.5 Concluding remarks

Science funding institutions are not mere means toward the ends of valuable science. Their functioning depends on values, only some of which concern the science itself, while others depend on procedural norms and how they fit into different socio-political nexuses. While further analysis and empirical study on values in science funding is needed, I hope to have taken steps toward understanding how values work in science funding practice, which can help clarify the values we should expect and promote moving forward (Figure 25.1).

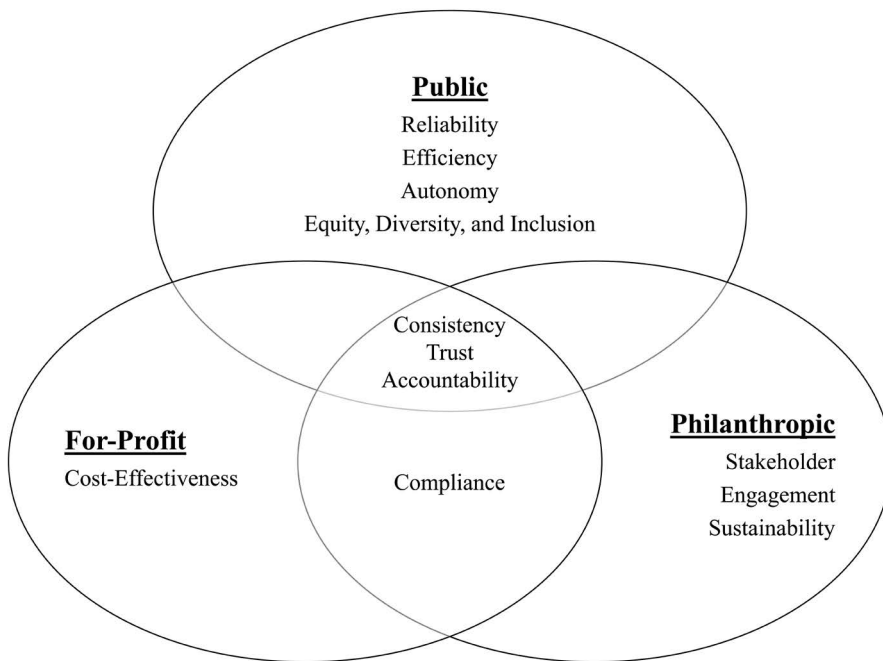


Figure 25.1 Values of Funding Agencies.

Notes

- 1 <https://cdn.pfizer.com/pfizercom/Clinical-Trial-Transparency-Policy-Paper-FINAL-2019.pdf>
- 2 <https://www.pkfeyerabend.org/en/>. It is, of course, no coincidence that a foundation named after Paul Feyerabend does not employ demarcation criteria between science and non-science. See Borrini-Feyerabend (2025) for further discussion.

References

- Aagaard, Kaare, Alexander Kladakis, and Mathias W. Nielsen. 2020. "Concentration or Dispersal of Research Funding?" *Quantitative Science Studies* 1 (1): 117–49. https://doi.org/10.1162/qss_a_00002.
- Aagaard, Kaare, Philippe Mongeon, Irene Ramos-Vielba, and Duncan Andrew Thomas. 2021. "Getting to the Bottom of Research Funding: Acknowledging the Complexity of Funding Dynamics." *PLoS One* 16 (5): e0251488. <https://doi.org/10.1371/journal.pone.0251488>.
- Adewoye, Deji, John Mendy, Emeka Smart Oruh, Chima Mordi, Arthur Egwuonwu, and Olutayo Otubanjo. 2023. "Africapitalism: The Marketisation of Philanthrocapitalism and Neoliberalism in African Entrepreneurial Philanthropy." *Marketing Theory* 23 (4): 685–708. <https://doi.org/10.1177/14705931231190954>.
- Aerts, Céline, Temmy Sunyoto, Fabrizio Tediosi, and Elisa Sicuri. 2017. "Are Public-Private Partnerships the Solution to Tackle Neglected Tropical Diseases? A Systematic Review of the Literature." *Health Policy* 121 (7): 745–54. <https://doi.org/10.1016/j.healthpol.2017.05.005>.
- Akomea-Frimpong, Isaac, Xiaohua Jin, Robert Osei-Kyei, and Roksana Jahan Tumpa. 2023. "A Critical Review of Public-Private Partnerships in the COVID-19 Pandemic: Key Themes and Future Research Agenda." *Smart and Sustainable Built Environment* 12 (4): 701–20. <https://doi.org/10.1108/SASBE-01-2022-0009>.
- Arco-Castro, Maria Lourdes, Maria Victoria Lopez-Pérez, Sara Rodriguez-Gomez, and Raquel Garde-Sánchez. 2020. "Do Stakeholders Modulate Philanthropic Strategy? Corporate Philanthropy as Stakeholders' Engagement." *Sustainability* 12 (18): 7242. <https://doi.org/10.3390/su12187242>.
- Aronson, Jeffrey K., and A. Richard Green. 2020. "Me-Too Pharmaceutical Products: History, Definitions, Examples, and Relevance to Drug Shortages and Essential Medicines Lists." *British Journal of Clinical Pharmacology* 86 (11): 2114–22. <https://doi.org/10.1111/bcp.14327>.
- Backer, Thomas E., Ralph Smith, and Ira Barbell. 2005. "Who Comes to the Table? Stakeholder Interactions in Philanthropy." *Rethinking Philanthropic Effectiveness*, edited by Dirk Elinghoff. Verlag Bertelsmann Stiftung.
- Baldwin, Melinda. 2018. "Scientific Autonomy, Public Accountability, and the Rise of 'Peer Review' in the Cold War United States." *Isis* 109 (3): 538–58. <https://doi.org/10.1086/700070>.
- Barnett, Michael. 2023. "Neoliberalism, Philanthropy, and Humanitarianism: Pragmatic or Faustian Bargains?" In *The Routledge International Handbook of Critical Philanthropy and Humanitarianism*, edited by Katharine Mitchell and Polly Pallister-Wilkins. Routledge.
- Benson-Greenwald, Tessa M., Alejandro Trujillo, Andrew D. White, and Amanda B. Diekman. 2023. "Science for Others or the Self? Presumed Motives for Science Shape Public Trust in Science." *Personality and Social Psychology Bulletin* 49 (3): 344–60. <https://doi.org/10.1177/01461672211064456>.
- Bloch, Carter, Irene Ramos-Vielba, Duncan Thomas, et al. 2023. "Promoting the Societal Impact of Research Funding." In *27th International Conference on Science, Technology and Innovation Indicators*.
- Borrini-Feyerabend, G. 2025. "The Paul K. Feyerabend Foundation: In Line with Paul's Work?." *Transversal: International Journal for the Historiography of Science* (19). <https://doi.org/10.24117/2526-2270.2025.i19.08>.
- Bravo, Pedro. 2026. "Values and Industry-Funded Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Bravo, Pedro, and Kevin C. Elliott. 2025. "The Promise and Perils of NGO-Supported Science." *Philosophy Compass* 20: e70043. <https://doi.org/10.1111/phc3.70043>.
- Brent, Thomas. 2024. "Poland Backs Embattled Science Funding Agency but Warns of Need for 'Transparency'." *Science Business*, April 16. https://sciencebusiness.net/news/eu-budget/poland-backs-embattled-science-funding-agency-warns-need-transparency?utm_source=ActiveCampaign&utm_medium=email&utm_content=EU+science+advisers+back+call+for+a++CERN+for+AI++to+aid+research&utm_campaign=Science%7CBusiness+Bulletin+No++1204.
- Chalmers, Don, and Dianne Nicol. 2004. "Commercialisation of Biotechnology: Public Trust and Research." *International Journal of Biotechnology* 6 (2–3): 116–33. <https://doi.org/10.1504/IJBT.2004.004806>.
- Chaudhry, Suparna, Marc Dotson, and Andrew Heiss. 2021. "Who Cares About Crackdowns? Exploring the Role of Trust in Individual Philanthropy." *Global Policy* 12: 45–58. <https://doi.org/10.1111/1758-5899.12984>.

- Collins, H. M. 2000. "Surviving Closure: Post-Rejection Adaptation and Plurality in Science." *American Sociological Review* 65 (6): 824–45. <https://doi.org/10.1177/000312240006500602>.
- Cozzens, Susan E. 1999. "Are New Accountability Rules Bad for Science?" *Issues in Science and Technology* 15 (4): 59–66. <https://www.jstor.org/stable/43313963>.
- Critchley, Christine, Dianne Nicol, and Margaret Otlowski. 2015. "The Impact of Commercialisation and Genetic Data Sharing Arrangements on Public Trust and the Intention to Participate in Biobank Research." *Public Health Genomics* 18 (3): 160–72. <https://doi.org/10.1159/000375441>.
- Cunningham, James A., Vincent Mangematin, Conor O’Kane, and Paul O’Reilly. 2016. "At the Frontiers of Scientific Advancement: The Factors that Influence Scientists to Become or Choose to Become Publicly Funded Principal Investigators." *The Journal of Technology Transfer* 41: 778–97. <https://doi.org/10.1007/s10961-015-9400-4>.
- Diamond, Arthur M. 1999. "Does Federal Funding ‘Crowd In’ Private Funding of Science?" *Contemporary Economic Policy* 17 (4): 423–31. <https://doi.org/10.1111/j.1465-7287.1999.tb00694.x>.
- Diamond, Arthur M. 2006. "The Relative Success of Private Funders and Government Funders in Funding Important Science." *European Journal of Law and Economics* 21: 149–61. <https://doi.org/10.1007/s10657-006-6647-0>.
- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2018. "From Tapestry to Loom: Broadening the Perspective on Values in Science." *Philosophy, Theory, and Practice in Biology* 10 (8): 1–8. <https://doi.org/10.3998/ptpbio.16039257.0010.008>.
- Douglas, Heather. 2024. "Structuring Institutions for Responsible and Accountable Science." *Philosophy of Science* 91 (5): 1149–58. <https://doi.org/10.1017/psa.2023.128>.
- Dresler, Martin. 2023. "FENS-Kavli Network of Excellence: Postponed, Non-Competitive Peer Review for Research Funding." *European Journal of Neuroscience* 58 (12): 4441–8. <https://doi.org/10.1111/ejn.15818>.
- Edwards, Marc A., and Siddhartha Roy. 2017. "Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition." *Environmental Engineering Science* 34 (1): 51–61. <https://doi.org/10.1089/ees.2016.0223>.
- Elliott, Kevin C. 2011. *Is a Little Pollution Good for You?: Incorporating Societal Values in Environmental Research*. Oxford University Press.
- Elliott, Kevin C. 2014. "Financial Conflicts of Interest and Criteria for Research Credibility." *Erkenntnis* 79 (Suppl 5): 917–37. <https://doi.org/10.1007/s10670-013-9536-2>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C., and Daniel J. McKaughan. 2009. "How Values in Scientific Discovery and Pursuit Alter Theory Appraisal." *Philosophy of Science* 76 (5): 598–611. <https://www.jstor.org/stable/10.1086/605807>.
- Epstein, Steven. 1996. *Impure Science: AIDS, Activism, and the Politics of Knowledge*. University of California Press.
- Etzkowitz, Henry, and Marina Ranga. 2011. "Gender Dynamics in Science and Technology: From the ‘Leaky Pipeline’ to the ‘Vanish Box.’" *Brussels Economic Review* 54 (2/3): 131–47.
- Evans, James A., Jae-Mahn Shim, and John P. A. Ioannidis. 2014. "Attention to Local Health Burden and the Global Disparity of Health Research." *PLoS One* 9 (4): e90147. <https://doi.org/10.1371/journal.pone.0090147>.
- Fang, Ferric C., and Arturo Casadevall. 2016. "Research Funding: The Case for a Modified Lottery." *mBio* 7 (2): 10–1128. <https://doi.org/10.1128/mbio.00422-16>.
- Feller, Irwin. 2022. "Assessing the Societal Impact of Publicly Funded Research." *The Journal of Technology Transfer* 47 (3): 632–50. <https://doi.org/10.1007/s10961-017-9602-z>.
- Fernández Pinto, Manuela. 2020. "Commercial Interests and the Erosion of Trust in Science." *Philosophy of Science* 87 (5): 1003–13. <https://doi.org/10.1086/710521>.
- Feyerabend, Paul. 1991. "Concluding Unphilosophical Conversation." In *Beyond Reason: Essays on the Philosophy of Paul K. Feyerabend*, edited by Gonzalo Munévar. Springer.
- Guthrie, Susan, Ioana Ghiga, and Steven Wooding. 2017. "What Do We Know About Grant Peer Review in the Health Sciences?" *F1000Research* 6. <https://doi.org/10.12688/f1000research.11917.2>.
- Hall, Kara L., Amanda L. Vogel, Grace C. Huang, et al. 2018. "The Science of Team Science: A Review of the Empirical Evidence and Research Gaps on Collaboration in Science." *American Psychologist* 73 (4): 532. <https://doi.org/10.1037/amp0000319>.
- Harvey, Charles, Jillian Gordon, and Mairi Maclean. 2021. "The Ethics of Entrepreneurial Philanthropy." *Journal of Business Ethics* 171: 33–49. <https://doi.org/10.1007/s10551-020-04468-7>.
- Herbert, Danielle L., Adrian G. Barnett, Philip Clarke, and Nicholas Graves. 2013. "On the Time Spent Preparing Grant Proposals: An Observational Study of Australian Researchers." *BMJ Open* 3 (5): e002800. <https://doi.org/10.1136/bmjopen-2013-002800>.

- Herder, Matthew. 2008. "Asking for Money Back-Chilling Commercialization or Recouping Public Trust in the Context of Stem Cell Research?" *Columbia Science and Technology Law Review* 9: 203.
- Heshmati, Ata, and Seyed Mehran Dibaji. 2019. "Science, Technology, and Innovation Status in Iran: Main Challenges." *Science, Technology and Society* 24 (3): 545–78. <https://doi.org/10.1177/0971721819873192>.
- Hipólito, Juliana, Leila Teruko Shirai, Luisa Maria Diele-Viegas, Rosana Halinski, Sílvia Soares Pires Carmen, and Gouveia Fontes Eliana Maria. 2022. "Brazilian Budget Cuts Further Threaten Gender Equality in Research." *Nature Ecology & Evolution* 6 (3): 234. <https://doi.org/10.1038/s41559-021-01640-8>.
- Hofstra, Bas, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Brian He, Dan Jurafsky, and Daniel A. McFarland. 2020. "The Diversity–Innovation Paradox in Science." *Proceedings of the National Academy of Sciences* 117 (17): 9284–91. <https://doi.org/10.1073/pnas.1915378117>.
- Holbrook, J. Britt. 2005. "Assessing the Science–Society Relation: The Case of the US National Science Foundation's Second Merit Review Criterion." *Technology in Society* 27 (4): 437–51. <https://doi.org/10.1016/j.techsoc.2005.08.001>.
- Hong, Jin, Song Hong, Lianbing Wang, Yi Xu, and Dingtao Zhao. 2015. "Government Grants, Private R&D Funding and Innovation Efficiency in Transition Economy." *Technology Analysis & Strategic Management* 27 (9): 1068–96. <https://doi.org/10.1080/09537325.2015.1060310>.
- Horbach, Serge P. J. M., Joeri K. Tjink, and Lex Bouter. 2022. "Research Funders Should be More Transparent: A Plea for Open Applications." *Royal Society Open Science* 9 (10): 220750. <https://doi.org/10.1098/rsos.220750>.
- Isaacs, David. 2016. "Work-Life Balance." *Journal of Paediatrics and Child Health* 52 (1): 5–6. <https://doi.org/10.1111/jpc.13110>.
- Jain, Monika. 2012. "Philanthro-Capitalism: Re-Modeling the Philanthropic Paradigm." *Decision* 39 (2): 45–61.
- Juk, Yohanna, Sergio Salles-Filho, Karen E. F. Pinto, et al. 2023. "Diversity, Equity and Inclusion: How Funding Agencies Are Addressing Inequalities in Research." In *27th International Conference on Science, Technology and Innovation Indicators (STI 2023)*.
- Kanji, Salmaan. 2015. "Turning Your Research Idea into a Proposal Worth Funding." *The Canadian Journal of Hospital Pharmacy* 68 (6): 458. <https://doi.org/10.4212/cjhp.v68i6.1502>.
- Kealey, Terence, and Omar Al-Ubaydli. 2000. "Should Governments Fund Science?" *Economic Affairs* 20 (3): 4–9. <https://doi.org/10.1111/1468-0270.00231>.
- Kevles, Daniel K. 1992. "Foundations, Universities, and Trends in Support for the Physical and Biological Sciences, 1900–1992." *Daedalus* 121 (4): 195–235. <https://www.jstor.org/stable/20027143>.
- Kozlov, Max. 2022. "NIH Plans Overhaul of Grant Review to Reduce Bias." *Nature News*, December 9. <https://doi.org/10.1038/d41586-022-04385-x>.
- Krimsky, Sheldon. 1999. "The Profit of Scientific Discovery and Its Normative Implications." *Chicago-Kent Law Review* 75: 15.
- Legg, Tess, Jenny Hatchard, and Anna B. Gilmore. 2021. "The Science for Profit Model—How and Why Corporations Influence Science and the Use of Science in Policy and Practice." *PLoS One* 16 (6): e0253272. <https://doi.org/10.1371/journal.pone.0253272>.
- Lepori, Benedetto. 2011. "Coordination Modes in Public Funding Systems." *Research Policy* 40 (3): 355–67. <https://doi.org/10.1016/j.respol.2010.10.016>.
- Lindner, Mark D, and Richard K. Nakamura. 2015. "Examining the Predictive Validity of NIH Peer Review Scores." *PLoS One* 10 (6): e0126938. <https://doi.org/10.1371/journal.pone.0132202>.
- Maienschein, Jane, Mary Sunderland, Rachel A. Ankeny, and Jason Scott Robert. 2008. "The Ethos and Ethics of Translational Research." *The American Journal of Bioethics* 8 (3): 43–51. <https://doi.org/10.1080/15265160802109314>.
- Mallapaty, Smriti. 2022. "What Japan's Election Means for Controversial Defence Research." *Nature News*, July 22. <https://www.nature.com/articles/d41586-022-02017-y>.
- Manning, Suzanne, and Matthew Walton. 2021. "Public Trust in Science and Research: Responsibility and Ethics." *Policy Quarterly* 17 (1). <https://doi.org/10.26686/pq.v17i1.6727>.
- Martino, Joseph. 2017. *Science Funding: Politics and Porkbarrel*. Routledge.
- Mati, Jacob Mwathi. 2017. "Philanthropy in Contemporary Africa: A Review." *Voluntaristics Review* 1 (6): 1–100. <https://doi.org/10.1163/24054933-12340014>.
- McGinn, Michelle K., Sandra Acker, Marie Vander Kloet, and Anne Wagner. 2019. "Dear SSHRC, What do you Want? An Epistolary Narrative of Expertise, Identity, and Time in Grant Writing." *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 20 (1): 1–24. <https://doi.org/10.17169/fqs-20.1.3128>.
- McGoey, Linsey. 2021. "Philanthrocapitalism and the Separation of Powers." *Annual Review of Law and Social Science* 17 (1): 391–409. <https://doi.org/10.1146/annurev-lawsocsci-120220-074323>.

- Michelson, Evan S. 2021. "Science Philanthropy, Energy Systems Research, and Societal Responsibility: A Match Made for the 21st Century." *Energy Research & Social Science* 72: 101886. <https://doi.org/10.1016/j.erss.2020.101886>.
- Mukerji, C. 1990. *A Fragile Power: Science and the State*. Princeton University Books.
- Myers, Kyle. 2020. "The Elasticity of Science." *American Economic Journal: Applied Economics* 12 (4): 103–34. <https://doi.org/10.1257/app.20180518>.
- Myers, Kyle, and Wei Yang Tham. 2023. "Money, Time, and Grant Design." *arXiv preprint arXiv: 2312.06479*. <https://doi.org/10.48550/arXiv.2312.06479>.
- Nishtar, Sania. 2004. "Public-Private 'Partnerships' in Health—A Global Call to Action." *Health Research Policy and Systems* 2: 1–7. <https://doi.org/10.1186/1478-4505-2-5>.
- O'Grady, Cathleen. 2022. "Upheaval in Norwegian Science Funding Threatens Grants." *Science News*, June 1. <https://www.science.org/content/article/upheaval-norwegianscience-funding-threatens-grants>.
- Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24 (sup1): 21–42. <https://doi.org/10.1080/00455091.1994.10717393>.
- Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Pamuk, Zeynep. 2019. "The British Academy Brian Barry Prize Essay Justifying Public Funding for Science." *British Journal of Political Science* 49 (1): 1–16. <https://doi.org/10.1017/S0007123418000431>.
- Plottel, Gloria Stone, Rachel Adler, Chelsea Jenter, and Jason P. Block. 2020. "Managing Conflicts and Maximizing Transparency in Industry-Funded Research." *AJOB Empirical Bioethics* 11 (4): 223–32. <https://doi.org/10.1080/23294515.2020.1798562>.
- Podhora, Aranka, Katharina Helming, Lucie Adenäuer, et al. 2013. "The Policy-Relevancy of Impact Assessment Tools: Evaluating Nine Years of European Research Funding." *Environmental Science & Policy* 31: 85–95. <https://doi.org/10.1016/j.envsci.2013.03.002>.
- Pulkkinen, Karoliina. 2026. "Values in Science in Nondemocratic Contexts." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Reiss, Julian, and Philip Kitcher. 2009. "Biomedical Research, Neglected Diseases, and Well-Ordered Science." *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia* 24 (3): 263–82.
- Sato, Sayaka, Pascal Mark Gyga, Julian Randall, and Marianne Schmid Mast. 2021. "The Leaky Pipeline in Research Grant Peer Review and Funding Decisions: Challenges and Future Directions." *Higher Education* 82 (1): 145–62. <https://doi.org/10.1007/s10734-020-00626-y>.
- Sismondo, Sergio. 2018. *Ghost Managed Medicine: Big Pharma's Invisible Hands*. Mattering Press.
- Sismondo, Sergio. 2021. "Epistemic Corruption, the Pharmaceutical Industry, and the Body of Medical Science." *Frontiers in Research Metrics and Analytics* 6: 614013. <https://doi.org/10.3389/frma.2021.614013>.
- Sismondo, Sergio, and Zdenka Chloubova. 2016. "'You're Not Just a Paid Monkey Reading Slides': How Key Opinion Leaders Explain and Justify Their Work." *BioSocieties* 11: 199–219. <https://doi.org/10.1057/biosoc.2015.32>.
- Solovey, Mark. 2020. *Social Science for What?: Battles over Public Funding for the "Other Sciences" at the National Science Foundation*. MIT Press.
- Steinberg, Alan. 2011. "Space Policy Responsiveness: The Relationship Between Public Opinion and NASA Funding." *Space Policy* 27 (4): 240–6. <https://doi.org/10.1016/j.spacepol.2011.07.003>.
- Stevens, Kelly R., Kristyn S. Masters, P. I. Imoukhuede, et al. 2021. "Fund Black Scientists." *Cell* 184 (3): 561–5.
- Taffe, Michael A., and Nicholas W. Gilpin. 2021. "Racial Inequity in Grant Funding from the US National Institutes of Health." *eLife* 10: e65697. <https://doi.org/10.7554/eLife.65697>.
- Toomey, Anne Helen. 2024. *Science with Impact: How to Engage People, Change Practice, and Influence Policy*. Island Press.
- von Hippel, Ted, and Courtney von Hippel. 2015. "To Apply or Not to Apply: A Survey Analysis of Grant Writing Costs and Benefits." *PLoS One* 10 (3): e0118494. <https://doi.org/10.1371/journal.pone.0118494>.
- Weingart, Peter. 1999. "Scientific Expertise and Political Accountability: Paradoxes of Science in Politics." *Science and Public Policy* 26 (3): 151–61. <https://doi.org/10.3152/147154399781782437>.
- Weinryb, Noomi. 2020. "Accountable and Thus Legitimate? A Comparative Study of Philanthropists." *Nonprofit and Voluntary Sector Quarterly* 49 (6): 1210–32. <https://doi.org/10.1177/0899764020925916>.
- Willholt, Torsten. 2010. "Scientific Freedom: Its Grounds and Their Limitations." *Studies in History and Philosophy of Science Part A* 41 (2): 174–81. <https://doi.org/10.1016/j.shpsa.2010.03.003>.

VALUES AND INDUSTRY-FUNDED RESEARCH

Pedro Bravo

26.1 Introduction

Industry-funded research is an unavoidable feature of contemporary science. Around two-thirds of the funding for R&D in the United States and in the European Union comes from industry. Although it may be less prevalent in countries in the Global South, it also has important implications there ([de Oliveira 2021](#); [Fernández Pinto 2021](#)). Several factors account for the emergence of this scenario. During the twentieth century, events inspired by neoliberalism, such as legislation strengthening intellectual property rights or, more broadly, different roles for the industry (e.g., outsourcing of R&D) and the State (e.g., withdrawal from its role as a primary provider of education), played crucial roles ([Krimsky 2003](#); [Mirowski 2011](#)). This chapter addresses the potential positive and negative consequences of increased industry funding in science, specifically examining how financial values shape research and exploring strategies to manage or mitigate these influences.

Given the potentially controversial nature of this issue and the likelihood that it may have already raised concerns, a few clarifications are necessary before moving forward. First, it is important to highlight that the themes to be introduced are important even if one is not committed to the thesis that identifying a source of funding is necessary to evaluate how reliable a specific study is. As I will show, industry funding in science has consequences far beyond its reliability. More importantly, it need not involve biased scientists or blatant cases of scientific misconduct. Some of the problems relating to industry-funded science may involve subtle choices in the design of a study or communication practices that affect the body of literature ([Doucet and Sismondo 2008](#); [Jukola 2021](#)). Third, the philosophical and scientific literature about it has pointed out how different kinds and degrees of industry influence may have different consequences. Industry funding in academic-industry partnerships may positively impact publication outcomes, but only up to a certain proportion of projects. Beyond that point, a larger share of industry involvement can have a negative effect ([Banal-Estañol, Jofre-Bonet, and Lawson 2015](#)). Industry-problematic involvement in studies may be more likely when certain conditions are met, such as the presence of ambiguity or uncertainty in the scientific findings and high financial incentives to influence the findings ([Elliott 2014](#)). Furthermore, some of the strategies used in widely known scandals (e.g., creating sympathetic scientific panels) are very resource-intensive, so it is not something that any corporation can do ([McGarity and Wagner 2008](#)).

Notwithstanding, one should not downplay the importance of the negative consequences of industry-funded science. Not only can companies impact the public trust in science, but they can

also lead to significant social and environmental harms (Fernández Pinto 2020a; Michaels 2020; Shrader-Frechette 2018). By drawing upon industry documents disclosed during litigations, formal methods, or case studies, scholars who investigate these issues can also shed some light on how to overcome them.

With that in mind, this chapter is structured as follows. The next three sections are devoted to stages in science where choices can favor some values over others: Agenda setting (Section 26.2), design and conduct of a study and the use of scientific categories (Section 26.3), and communication of scientific findings (Section 26.4). I provide an overview of the positive and negative consequences of industry-funded science in each of these stages and the industry's influence on the social context that bears on them. Thus, it will be possible to see industry impact both from a macro and a micro level of analysis. In Section 26.5, I list some of the solutions proposed to the drawbacks of industry-funded science, dividing them into two categories: More transparency and more independence. Finally, I conclude with some suggestions for further research.

26.2 Agenda setting

Industry-funded research is commonly praised thanks to its application orientation (Caulfield and Ogbogu 2015; Drazen 2015). Take, for example, the list of devices developed by Bell Labs: The laser, the transistor, important programming languages, etc. In other words, commercial interests driving research can help develop theoretical ideas into technological and profitable solutions to practical problems. Without these commercial interests, one can argue that part of the current research done in fields such as physics, chemistry, and biology would not be done (Carrier 2008, 224). Therefore, it is possible to state that industry funding broadens, in this sense, the questions investigated by scientists.

Although this characterization may be common, it might stimulate simplistic ideas about the practice of industry application-oriented research, such as that it involves the mere application of previous ideas into new applications. As some authors have been highlighting, this kind of research may both improve the theoretical knowledge necessary to develop a technology and generate a kind of knowledge closely related to the particularities of the developed object (Wilholt 2006; see also de Vries 2011).

Despite these advantages, there are at least three relevant drawbacks on the other side of the coin. First, themes other than those related to producing profitable goods and services are understudied (Fabbri et al. 2018). More specifically, in the biomedical field, not only do rare and tropical diseases get less attention, but, more worryingly, some authors argue that certain health solutions are only studied because of their profitability and not because of medical benefits (Brown 2017).

Second, there is also here the phenomenon of funding *friendly research* or *distraction science* (Fernández Pinto 2017; Galison and Proctor 2020, 29). This phenomenon generally involves funding studies that search for alternative explanations about associations between a product and a harmful effect to manufacture doubt and then delay or avoid regulations. A classic example is the tobacco industry funding studies about several other factors that could account for the association between smoking and cancer. A more recent one is Coca-Cola's funding of studies about the effects of exercise on obesity to downplay the role of calorie consumption (Nestle 2018). This phenomenon is interesting because it does not imply that the science or the scientists involved in those studies were necessarily biased. For example, one of the earlier studies funded by the tobacco industry rightly established the association between asbestos and lung cancer (Oreskes and Conway 2010, 17). Additionally, the industry may choose to fund and amplify the research of scientists whose previous views were already sympathetic to the industry's commercial interests, thus impacting the body of literature without changing individual scientists' perspectives (Holman and Bruner 2017).

Third, industry can impact the social context of science in a way that delays or prevents some lines of inquiry. One of the reasons for this is the failure to share industry studies, a phenomenon that will be discussed in [Section 26.4](#). For now, it suffices to state that this non-release of information can delay relevant studies for decades. For example, 3M had internal studies since 1978 about per- and polyfluoroalkyl substances' (PFAS) effects on the immune system, but they were not published. What Philippe Grandjean, the researcher responsible for investigating immunological effects of PFAS in the 2000s, stated about the situation is very illuminating:

Had I found out in 1978 that this industrial chemical was toxic to the immune system, I could see all sorts of examinations of exposed kids that could be done, but I was not told, so it had to wait, [in] this case 30 years, before I turned my attention to this.

(Grandjean as cited in Michaels 2020)

The second reason has to do with how industry groups or organizations heavily supported by them can lobby public agencies and politicians to prevent the study of certain objects. For example, the National Rifle Association in the US influenced a congressman to add an amendment to a federal spending bill that blocked the Centers for Disease Control and Prevention (CDC) from funding research on gun violence ([Dzau and Leshner 2018](#)).

By preventing or selectively choosing to fund some themes over others, the industry can contribute to the social construction of ignorance, a theme investigated by the field called *agnotology* ([Kourany and Carrier 2020](#)). If the industry has its research priorities and its proportion of scientific funding is increasing, one can legitimately worry that important themes would not be investigated, thus resulting in a form of ignorance as a by-product of the preferred lines of inquiry. Another way industries can foster ignorance is by influencing policymakers to block specific lines of inquiry, as illustrated by the National Rifle Association example earlier. This can also involve tactics like intimidating scientists (as discussed by [Fernández Pinto and Leuschner 2026](#) and [Jacquet 2022](#)).

26.3 Design and conduct of a study and the use of scientific categories

There are several cases in which the design of a study or the analysis of data was made in a way to favor a pro-industry outcome. This need not always be the case, however. For some authors, the same commercial interests driving industry-funded science to be directed toward some themes rather than others can also be an important motivation for doing science in ways that generate reliability. Delivering a technological good that fails may break an industry; similarly, at least in principle, competition among companies could also be an important factor for stimulating one company to expose the failures of others and then take advantage of them ([Carrier 2008](#), 225–26). Industry labs following standardized protocols and Good Laboratory Practices (a system for ensuring reliable record-keeping in industry science performed for regulatory purposes) could also be an important factor driving reliability ([Conrad and Becker 2011](#)). In a period with concerns regarding a potential reproducibility crisis, partnerships with industry are often defended as a possible solution ([Edwards 2016](#)).

These and other features of industry-funded research are sometimes invoked when discussing the phenomenon of the funding effect. The *funding effect* is the systematic tendency of industry-sponsored research to lead to pro-industry outcomes in comparison to similar studies funded by the government or non-governmental organizations. It was first uncovered in pharmaceutical studies, but later in other areas as diverse as nutrition ([Jukola 2021](#)), indoor tanning ([Adekunle et al. 2020](#)), and chemical safety studies ([Volz and Elliott 2012](#)). A recent example is provided by [Xie and Zhou \(2022\)](#), who found that industry-funded studies of new treatments were more likely to find them to be cost-effective than non-industry-sponsored studies.

Some hypothetical causes for the funding effect do not involve epistemically inappropriate decisions. For example, some authors argue that it can result from larger trials that the industry can fund and, consequently, a higher statistical power in detecting differences among groups. Another explanation would be that the industry may have preliminary positive data about one of its products, which stimulates additional studies on it. Furthermore, there is evidence that industry-sponsored studies tend to be better concerning blinding than non-industry-sponsored studies (Lundh et al. 2017, 18). For reasons like these, some authors defend that the funding effect is only a *prima facie* reason to suspect the presence of bias and not a sufficient one for it (Krimsky 2013; see also de Melo-Martín 2019).

Notwithstanding, one of the most recent meta-analyses about the funding effect suggests that “it seems most plausible that industry achieves overly positive results through a variety of biasing choices in the design, conduct and reporting of their studies” (Lundh et al. 2017, 19). If so, there is no funding effect, but rather *industry bias*. Lundh et al.’s analysis supported this conclusion by controlling for the sample size and other confounders and showing that the funding effect was still present. Another reason they provide to support their claim is a comparison of studies sponsored by the manufacturer of the test treatment with those sponsored by the manufacturer of the control treatment: Trials funded by the manufacturer of the test treatment were much more likely to favor it than when funded by the manufacturer of the control treatment (Lundh et al. 2017, 13).

The choices that contribute to the funding effect may be various. In the next section, I will mention choices related to the reporting of studies, such as selective reporting of outcomes. Choices in the conduct of research may involve cases of fabrication of data, although it may not be as frequent or important as choices in the design of a study. In the pharmaceutical domain, choices in the design involve testing a drug against a treatment that either does not work or does not work very well, against a too-low dose of the comparison drug because it will make the drug appear more effective, or against too high a dose of the comparison drug because it will make the drug appear less toxic. In chemical safety studies, it may involve the use of less sensitive animals (Vom Saal et al. 2005, 247), the investigation of the harmful effects of individual substances instead of their combination with other chemicals (Krimsky 2021), or a short duration of the study. For example, Fagan, Antoniou, and Robinson (2014) contrast the duration (two years) of a study that found health hazards related to genetically modified corn by Monsanto with the company’s study (three months); since tumors in the rats only appeared after four months, Monsanto’s study was not able to document it (see also McHenry 2018). In a meta-analysis, it can involve selective sampling from the literature among other choices (Stegenga 2018, 88–95).

Regardless of industry bias, another worry about industry-funded science related to the design and conduct of studies has to do with the creation and use of categories or concepts. The values in science literature have been showing how values are necessary to define some scientific concepts and/or to analyze the social ramifications of adopting one concept instead of another (see the conceptual argument in Elliott’s chapter in this volume; Elliott 2026). In industry-funded science, though, some argue that the commercial values behind it illegitimately influence some concepts. An example of the first kind of influence of values is the phenomenon of disease-mongering, where the extensions of diseases are broadened so that pharmaceutical companies can have a larger number of consumers for their drugs (González-Moreno et al. 2015). Another example from environmental science is the livestock industry’s promotion of “climate neutrality” as having no additional warming impact, rather than its original definition of net-zero greenhouse gas emissions. This shift in meaning allows a livestock sector with high but decreasing methane emissions to still claim climate neutrality (Donnison and Murphy-Bokern 2023). Finally, an example of the second kind of influence is the suggestion of a 1996 conference where many researchers associated with the chemical industry replaced the term “multiple chemical sensitivity” with “idiopathic environmental intolerance” to lessen the attention to chemicals as a cause of some people’s health problems (Elliott 2017, 126).

Finally, another important design choice comes from how to recruit patients for clinical trials in countries where there are weaker regulatory frameworks. For example, [Fernández Pinto \(2019, 399\)](#) details how the practice of Contract Research Organizations (CROs) for recruiting women and underrepresented minorities in Latin America has both ethical and epistemic consequences. Among them, the CROs' practice of paying doctors to stimulate patient recruitment can allow the inclusion of patients who do not meet the criteria of the study, which can affect the quality of the knowledge generated and impact their health (see also [Shah 2006](#), 141–62, for similar worries in India).

26.4 Communicating scientific findings

When focusing on decisions about how to communicate scientific findings, most of the literature about industry-funded science has been paying attention to its worrying elements. This should come as no surprise since a significant part of the industry's negative influence on research occurs through scientific communication. There may, however, be some benefits. For example, according to research by [Banal-Estañol, Jofre-Bonet, and Lawson \(2015\)](#), the relationship between industry collaborations and the publications of university scientists has the form of an inverted U. Therefore, up to some threshold, collaboration with industry “can boost research output,” an effect that does not hold with higher degrees of collaboration when several constraints take place ([Banal-Estañol, Jofre-Bonet, and Lawson 2015](#)). Notwithstanding, one could object that if the published studies are biased, the simple increase in publications would not count as a good consequence.

In general, though, the most common complaint about industry-funded science and science communication is how it conflicts with the norm according to which scientific findings should belong to everybody, a norm proposed by the sociologist of science Robert [Merton \(1942\)](#) and known as communism ([Krimsky 2003](#); [Resnik 2010](#)). Along with other norms proposed by Merton, communism is highly supported by scientists, and a recent survey of 633 scientists from South Korea showed that it was the most respected norm and the one believed to be most in tension with the increasing commercialization of science ([Bright and Heesen 2023](#), 6). [Blumenthal et al. \(2006\)](#) found evidence of this tension in their study of 1,849 life scientists. Their data showed an association between industry-funded research and the withholding of publications.

To see how industry-funded research conflicts with communism, one must just note how unwelcome results from industry-funded research can either remain unpublished or be selectively reported for a wide range of reasons. An industry can withhold results for competitive reasons or to avoid regulations and lawsuits in case their products are harmful. The selective reporting of its results can also make a product appear safer and/or more effective than it is. Instruments such as patents, confidential business information, material transfer agreements, and clauses in funding contracts restricting or delaying publication also promote a culture of secrecy as opposed to communalism ([Mirowski 2011](#)).

It is important to recognize that although public-funded research is not completely communal, the incentives for sharing in this context are, except for military science, stronger than in industry-funded research ([Bright and Heesen 2023](#); [Holman and Elliott 2018](#)). Not surprisingly, open science initiatives have been mostly focused on public-funded research, with the business sector benefiting from it ([Fernández Pinto 2020b](#); [Mirowski 2018](#)). When the objects of study have less commercial importance, withholding may be less severe among industry-funded scientists ([Evans 2010](#)). However, the non-publication of unwelcome results and selective reporting of industry-funded science still have significant epistemic and social consequences.

Epistemically, as we saw in [Section 26.2](#), it can prevent or delay socially important lines of research. In addition, both publication choices negatively affect the overall scientific literature, producing false or misleading beliefs ([Steel 2018](#)). Take the case of publication bias. A famous study by [Turner et al. \(2008\)](#) about antidepressant drugs submitted for the Food and Drug Administration

(FDA)’s approval found that negative results either tended not to be published or to be published as positive, which resulted in the following situation: While a little less than half of the studies submitted to the FDA had positive results, nearly all the published literature appeared to be positive.

Selective reporting can take several forms, such as including only a subset of the analyzed data in the published study. An example of it was described by [Michaels \(2008, 150\)](#): A company did a 12-month trial to investigate whether a drug was associated with reducing ulcers; no ulcer-reducing benefit appeared in 12 months, but over 6 months, a small one did, so they published only the data over 6 months. Selective reporting can also happen through rhetorical moves that put the results in a more favorable light, such as the omission of negative information in the abstract or conclusion of a published paper, even if the negative information is present in the results section ([Chiu, Grundy, and Bero 2017](#)). This concerning practice may also extend to educational materials sponsored by industry ([Beder 1998](#); [Rust 2011](#)), or to marketing efforts aimed at entering new markets, even when the risks associated with the products are already well-known and subject to restrictions elsewhere ([Lerner 2016](#)). In addition, an industry-affiliated scientist working closely with policymakers can also exert significant influence by selectively reporting the portion of the scientific literature that supports the industry’s view. This strategy is interesting because it does not presuppose biased studies and it can draw as well upon public-funded—especially low-powered—research ([Weatherall, O’Connor, and Bruner 2020](#)).

Socially, these publications and reporting choices all delay preventive or precautionary actions when there are potential harms associated with the objects under investigation, which makes it even harder to mitigate them ([Harremoës et al. 2001](#)). In the pharmaceutical domain, it affects how prescriptions are created, eventually harming people with unnecessary or unsafe drugs ([Sismondo 2018](#)). Withholding of results is also a waste of resources since the public literature will not reflect the studies that have been performed, and researchers may even spend other important resources to conduct a study without the opportunity to rely upon a previous attempt.

The discussions above are related to whether and what to publish or communicate, but industry-funded research has also impacted how publishing occurs through *ghostwriting* and *ghostmanagement*. Ghostwriting refers to writing for someone else who is the named author. On its own, it may not be problematic. However, pharmaceutical companies use this strategy by inviting renowned professors to sign as authors of the papers written by the companies’ scientists; in these cases, ghostwriting aims to increase the credibility of a paper by relying upon the authority of the invited and allegedly independent scientists. As extensively detailed by [Sismondo \(2018\)](#), ghostwriting is a part of the larger phenomenon of ghost management, where multiple steps relating to an article are controlled by the company to shape its message following its marketing goals.

26.5 Potential solutions

The diversity of worries presented in the last section implies that hardly a single solution will correct it. Thus, several solutions—sometimes in combination—have been proposed by authors who investigate the negative impact of industry-funded science. They vary per the author’s diagnosis about where the problem is located and its severity, but also regarding the agents (e.g., journals, funding agencies, academic institutions) who will enact or are the object of the proposals. With that in mind, I try to list below a diverse range of approaches by grouping them under two categories: (1) More transparency and (2) more independence.

- 1 Several proposals to mitigate the negative impact of industry on science have to do with increasing transparency. Disclosure of conflicts of interest, rigorous standards for trial reporting, and trial registries are examples of this category. Despite some drawbacks associated with how people

perceive conflicts of interest and the obvious feature that disclosure is not the same as eliminating, disclosure can increase the accountability of those who negatively influence science, and it may also facilitate the critical interaction of the scientific community with studies funded by industry (Oreskes et al. 2015; but see de Melo-Martín and Intemann 2009). Critical interaction may be one of the few resources to deal with problems such as influences on scientific categories. Standards for trial reporting can discourage selective reporting and some kinds of fraud, but they also have important limitations as they do not exclude the possibility of bias, such as in the choice of comparators (Doucet and Sismondo 2008, 627). Trial registries are an important improvement since they can help identify publication bias and selective reporting; nonetheless, they do not address other sources of bias, and their effectiveness depends on the details of their implementation (Goldacre 2015).

- 2 Another strategy to deal with industry-funded science is to make science as independent as possible from industry by either lessening or eliminating the influence of commercial interests. Increasing public funding for socially relevant themes, thus having science in the public interest and not only public-funded science, would be an example of this approach with the potential to counterbalance industry influence on the choice of research topics (Krimsky 2003). However, as some of the findings I mentioned point out, even public-funded research can be co-opted by industry. Discouraging review articles of authors with conflicts of interest (e.g., the *New England Journal of Medicine*'s editorial policies) or forbidding the submission of papers by companies with a long-standing bad track record in influencing science (e.g., tobacco companies) are also examples of this category with important impact on the published literature. Changes in intellectual property policies can also be a significant remedy against problems related to withholding of data in university–industry partnerships (Sterckx 2011). A science court in which industry is only an interested part of a jury, where it would be counter-balanced by scientists advocating a contrary position, could lessen its overall effects as well (Biddle 2013). More commonly, though, authors propose that the industry should only give money to an independent third party that would plan and conduct studies (Krimsky 2003; Volz and Elliott 2012). Although it may be politically difficult to enact this approach, it may be one of the most promising to deal with the sorts of problems we encountered in the agenda setting, design, and communication of findings. However, the details of its implementation are also important; otherwise, it can be co-opted by industry (Nestle 2018; see also the asymmetric arms race framework for a theoretical anticipation of this sort of problem in Holman and Geislar (2018)). Finally, a stronger position would be to have only publicly funded research and to have patents immediately eliminated (Brown 2017).

26.6 Conclusion: Next steps

The overview of industry-funded science's potential positive and negative consequences provided in the last sections can shed some light on what the “next steps” might be for philosophers who are thinking about values and industry. On one hand, an important and persistent gap in the literature deserving further studies concerns under which specific conditions industry-funded science is beneficial. In possession of those conditions, one could find ways to promote them and, consequently, decrease the industry's negative influence. On the other hand, there are some sectors of industry with a negative influence on science that, to the best of my knowledge, have not received scrutiny by philosophers of science (e.g., social media companies; Oremus 2021). It would be interesting to compare their strategies with more traditional industry sectors. Future studies could also pay more attention to how public-funded science is also impacted by the incentives and pressure of industry. Finally, increased interaction between philosophers and other scholars, policymakers, citizens, and investigative journalists could be very helpful.

References

- Adekunle, Lola, Rebbecca Chen, Lily Morrison, et al. 2020. "Association Between Financial Links to Indoor Tanning Industry and Conclusions of Published Studies on Indoor Tanning: Systematic Review." *BMJ* 368. <https://doi.org/10.1136/bmj.m7>.
- Banal-Estañol, Albert, Mireia Jofre-Bonet, and Cornelia Lawson. 2015. "The Double-Edged Sword of Industry Collaboration: Evidence from Engineering Academics in the UK." *Research Policy* 44 (6): 1160–75. <https://doi.org/10.1016/j.respol.2015.02.006>.
- Beder, Sharon. 1998. "The Dioxin Controversy: Spilling Over into Schools." *Australian Science Teachers' Journal* 44 (4): 28–34.
- Biddle, Justin. 2013. "Institutionalizing Dissent: A Proposal for an Adversarial System of Pharmaceutical Research." *Kennedy Institute of Ethics Journal* 23 (4): 325–53. <https://doi.org/10.1353/ken.2013.0013>.
- Blumenthal, David, Eric G. Campbell, Manjusha Gokhale, et al. 2006. "Data Withholding in Genetics and the Other Life Sciences: Prevalences and Predictors." *Academic Medicine* 81 (2): 137–45.
- Bright, Liam Kofi, and Remco Heesen. 2023. "To Be Scientific Is to Be Communist." *Social Epistemology* 37 (3): 249–58. <https://doi.org/10.1080/02691728.2022.2156308>.
- Brown, James R. 2017. "Socializing Medical Research." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Carrier, Martin. 2008. "Science in the Grip of the Economy: On the Epistemic Impact of the Commercialization of Research." In *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*, edited by Martin Carrier, Don Howard, and Janet A. Kourany. University of Pittsburgh Press.
- Caulfield, Timothy, and Ubaka Ogbogu. 2015. "The Commercialization of University-Based Research: Balancing Risks and Benefits." *BMC Medical Ethics* 16 (1). <https://doi.org/10.1186/s12910-015-0064-2>.
- Chiu, Kellia, Quinn Grundy, and Lisa Bero. 2017. "'Spin' in Published Biomedical Literature: A Methodological Systematic Review." *PLoS Biology* 15 (9). <https://doi.org/10.1371/journal.pbio.2002173>.
- Conrad, James W., and Richard A. Becker. 2011. "Enhancing Credibility of Chemical Safety Studies: Emerging Consensus on Key Assessment Criteria." *Environmental Health Perspectives* 119 (6): 757–64. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3114808/>.
- de Melo-Martin, Inmaculada. 2019. "The Commercialization of the Biomedical Sciences: (Mis)Understanding Bias." *History and Philosophy of the Life Sciences* 41 (3): 1–17. <https://doi.org/10.1007/s40656-019-0274-x>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2009. "How Do Disclosure Policies Fail? Let Us Count the Ways." *The FASEB Journal* 23 (6): 1638–42. <https://doi.org/10.1096/fj.08-125963>.
- de Oliveira, Marcos Barbosa. 2021. "Innovationism North and South." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zihle Mncube, Luana Poliseli, and Luis Reyes-Galindo. Routledge. <https://doi.org/10.4324/9781003027140>.
- de Vries, Marc J. 2011. "Science in the Context of Industrial Application: The Case of the Philips Natuurkundig Laboratorium." In *Science in the Context of Application*, edited by Martin Carrier and Alfred Nordmann. Springer. <https://doi.org/10.1007/978-90-481-9051-5>.
- Donnison, Caspar L., and Donal Murphy-Bokern. 2023. "Are Climate Neutrality Claims in the Livestock Sector Too Good to Be True?" *Environmental Research Letters* 19 (1). <https://doi.org/10.1088/1748-9326/ad0f75>.
- Doucet, Mathieu, and Sergio Sismondo. 2008. "Evaluating Solutions to Sponsorship Bias." *Journal of Medical Ethics* 34 (8): 627–30. <https://doi.org/10.1136/jme.2007.022467>.
- Drazen, Jeffrey M. 2015. "Revisiting the Commercial-Academic Interface." *New England Journal of Medicine* 372 (19): 1853–54. <https://doi.org/10.1056/NEJMe1503623>.
- Dzau, Victor J., and Alan I. Leshner. 2018. "Public Health Research on Gun Violence: Long Overdue." *Annals of Internal Medicine* 168 (12): 876–77. <https://doi.org/10.7326/M18-0579>.
- Edwards, Aled. 2016. "Reproducibility: Team Up with Industry." *Nature* 531 (7594): 299–301. <https://doi.org/10.1038/531299a>.
- Elliott, Kevin C. 2014. "Financial Conflicts of Interest and Criteria for Research Credibility." *Erkenntnis* 79 (S5): 917–37. <https://doi.org/10.1007/s10670-013-9536-2>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2026. "Arguments against the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Evans, James A. 2010. "Industry Collaboration, Scientific Sharing, and the Dissemination of Knowledge." *Social Studies of Science* 40 (5): 757–91. <https://doi.org/10.1177/0306312710379931>.
- Fabbri, Alice, Alexandra Lai, Quinn Grundy, and Lisa Anne Bero. 2018. "The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review." *American Journal of Public Health* 108 (11): e9–16. <https://doi.org/10.2105/AJPH.2018.304677>.

- Fagan, John, Michael Antoniou, and Claire Robinson. 2014. "Busting the Big GMO Myths." In *The GMO Deception*, edited by Sheldon Krimsky and Jeremy Gruber. Skyhorse Publishing.
- Fernández Pinto, Manuela. 2017. "To Know or Better Not to: Agnotology and the Social Construction of Ignorance in Commercially Driven Research." *Science & Technology Studies* 30 (2): 53–72. <https://doi.org/10.23987/sts.61030>.
- Fernández Pinto, Manuela. 2019. "Doubly Disadvantaged: On the Recruitment of Diverse Subjects for Clinical Trials in Latin America." *Tapuya: Latin American Science, Technology and Society* 2 (1): 391–407. <https://doi.org/10.1080/25729861.2018.1556237>.
- Fernández Pinto, Manuela. 2020a. "Commercial Interests and the Erosion of Trust in Science." *Philosophy of Science* 87 (5): 1003–13. <https://doi.org/10.1086/710521>.
- Fernández Pinto, Manuela. 2020b. "Open Science for Private Interests? How the Logic of Open Science Contributes to the Commercialization of Research." *Frontiers in Research Metrics and Analytics* 5. <https://doi.org/10.3389/frma.2020.588331>.
- Fernández Pinto, Manuela. 2021. "Science and Industry Funding." In *Global Epistemologies and Philosophies of Science*, edited by David Ludwig, Inkeri Koskinen, Zinhle Mncube, Luana Poliseli, and Luis Reyes-Galindo. Routledge. <https://doi.org/10.4324/9781003027140>.
- Fernández Pinto, Manuela, and Anna Leuschner. 2026. "Epistemic Intimidation and Illegitimate Value-Influences in Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Galison, Peter, and Robert Proctor. 2020. "Agnotology in Action: A Dialogue." In *Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted*, edited by Janet A. Kourany and Martin Carrier. The MIT Press.
- Goldacre, Ben. 2015. "How to Get All Trials Reported: Audit, Better Data, and Individual Accountability." *PLoS Medicine* 12 (4). <https://doi.org/10.1371/journal.pmed.1001821>.
- González-Moreno, María, Cristian Saborido, and David Teira. 2015. "Disease-Mongering through Clinical Trials." *Studies in History and Philosophy of Biological and Biomedical Sciences* 51: 11–18. <https://doi.org/10.1016/j.shpsc.2015.02.007>.
- Harremoës, Poul, David Gee, Malcolm MacGarvin, et al., eds. 2001. *Late Lessons from Early Warnings: The Precautionary Principle, 1896–2000*. European Environment Agency.
- Holman, Bennett, and Justin Bruner. 2017. "Experimentation by Industrial Selection." *Philosophy of Science* 84 (5): 1008–19. <https://doi.org/10.1086/694037>.
- Holman, Bennett, and Kevin C. Elliott. 2018. "The Promise and Perils of Industry-Funded Science." *Philosophy Compass* 13 (11): e12544. <https://doi.org/10.1111/phc3.12544>.
- Holman, Bennett, and Sally Geislar. 2018. "Sex Drugs and Corporate Ventriloquism: How to Evaluate Science Policies Intended to Manage Industry-Funded Bias." *Philosophy of Science* 85 (5): 869–81. <https://doi.org/10.1086/699713>.
- Jacquet, Jennifer. 2022. *The Playbook: How to Deny Science, Sell Lies, and Make a Killing in the Corporate World*. Pantheon Books.
- Jukola, Saana. 2021. "Commercial Interests, Agenda Setting, and the Epistemic Trustworthiness of Nutrition Science." *Synthese* 198 (10): 2629–46. <https://doi.org/10.1007/s11229-019-02228-3>.
- Kourany, Janet A., and Martin Carrier, eds. 2020. *Science and the Production of Ignorance: When the Quest for Knowledge Is Thwarted*. The MIT Press.
- Krimsky, Sheldon. 2003. *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* Rowman & Littlefield Publishers.
- Krimsky, Sheldon. 2013. "Do Financial Conflicts of Interest Bias Research?: An Inquiry into the 'Funding Effect' Hypothesis." *Science, Technology, & Human Values* 38 (4): 566–87. <https://doi.org/10.1177/0162243912456271>.
- Krimsky, Sheldon. 2021. "Glyphosate-Based Herbicides and Public Health: Making Sense of the Science." *Journal of Agricultural and Environmental Ethics* 35 (1). <https://doi.org/10.1007/s10806-021-09874-z>.
- Lerner, Saron. 2016. "The Teflon Toxin Goes to China." *The Intercept*, September 15. <https://theintercept.com/2016/09/15/the-teflon-toxin-goes-to-china/>.
- Lundh, Andreas, Joel Lexchin, Barbara Mintzes, Jeppe B. Schroll, and Lisa Bero. 2017. "Industry Sponsorship and Research Outcome." *Cochrane Database of Systematic Reviews* 2017 (2). <https://doi.org/10.1002/14651858.MR000033.pub3>.
- McGarity, Thomas O., and Wendy E. Wagner. 2008. *Bending Science: How Special Interests Corrupt Public Health Research*. Harvard University Press.

- McHenry, Leemon B. 2018. "The Monsanto Papers: Poisoning the Scientific Well." *International Journal of Risk & Safety in Medicine* 29 (3–4): 193–205. <https://doi.org/10.3233/JRS-180028>.
- Merton, Robert K. 1942. "A Note on Science and Democracy." *Journal of Legal and Political Sociology* 1 (1–2): 115–26. <https://doi.org/10.1515/9783110375008-013>.
- Michaels, David. 2008. *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford University Press.
- Michaels, David. 2020. *The Triumph of Doubt: Dark Money and the Science of Deception*. Oxford University Press.
- Mirowski, Philip. 2011. *Science-Mart: Privatizing American Science*. Harvard University Press.
- Mirowski, Philip. 2018. "The Future(s) of Open Science." *Social Studies of Science* 48 (2): 171–203. <https://doi.org/10.1177/0306312718772086>.
- Nestle, Marion. 2018. *Unsavory Truth: How Food Companies Skew the Science of What We Eat*. Basic Books.
- Oremus, Will. 2021. "Facebook Keeps Researching Its Own Harms—And Burying the Findings." *The Washington Post*, September 16. <https://www.washingtonpost.com/technology/2021/09/16/facebook-files-internal-research-harms/>.
- Oreskes, Naomi, Daniel Carlat, Michael E. Mann, Paul D. Thacker, and Frederick S. vom Saal. 2015. "Viewpoint: Why Disclosure Matters." *Environmental Science & Technology* 49 (13): 7527–28. <https://doi.org/10.1021/acs.est.5b02726>.
- Oreskes, Naomi, and Erik M. Conway, E. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Resnik, David B. 2010. "Financial Interests and the Norms of Academic Science." In *The Commodification of Academic Research*, edited by Hans Radder. University of Pittsburgh Press.
- Rust, Susan. 2011. "Plastics Industry Edited California's Environmental Textbook." *The Mercury News*, August 19. <https://www.mercurynews.com/2011/08/19/plastics-industry-edited-californias-environmental-textbook/>.
- Shah, Sonia. 2006. *The Body Hunters: Testing New Drugs on the World's Poorest Patients*. The New Press.
- Shrader-Frechette, Kristin. 2018. "Normative Philosophy of Science: Responding to Special-Interest Science." *Axiomathes* 28 (6): 679–93. <https://doi.org/10.1007/s10516-018-9402-6>.
- Sismondo, Sergio. 2018. *Ghost-Managed Medicine: Big Pharma's Invisible Hands*. Mattering Press.
- Steel, Daniel. 2018. "If the Facts Were Not Untruths, Their Implications Were: Sponsorship Bias and Misleading Communication." *Kennedy Institute of Ethics Journal* 28 (2): 119–44. <https://doi.org/10.1353/ken.2018.0011>.
- Stegenga, Jacob. 2018. *Medical Nihilism*. Oxford University Press.
- Sterckx, Sigrid. 2011. "Patenting and Licensing of University Research: Promoting Innovation or Undermining Academic Values?" *Science and Engineering Ethics* 17 (1): 45–64. <https://doi.org/10.1007/s11948-009-9168-8>.
- Turner, Erick H., Annette M. Matthews, Eftihia Linardatos, Robert A. Tell, and Robert Rosenthal. 2008. "Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy." *New England Journal of Medicine* 358 (3): 252–60. <https://doi.org/10.1056/NEJMsa065779>.
- Volz, David C., and Kevin C. Elliott. 2012. "Mitigating Conflicts of Interest in Chemical Safety Testing." *Environmental Science & Technology* 46 (15): 7937–38. <https://doi.org/10.1021/es3028296>.
- Vom Saal, Fredrick S., Susan C. Nagel, Barry G. Timms, and Wade V. Welshons. 2005. "Implications for Human Health of the Extensive Bisphenol A Literature Showing Adverse Effects at Low Doses: A Response to Attempts to Mislead the Public." *Toxicology* 212 (2–3): 244–52. <https://doi.org/10.1016/j.tox.2005.05.006>.
- Weatherall, James Owen, Cailin O'Connor, and Justin P. Bruner. 2020. "How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks." *The British Journal for the Philosophy of Science* 71 (4): 1157–86. <https://doi.org/10.1093/bjps/axy062>.
- Wilholt, Torsten. 2006. "Design Rules: Industrial Research and Epistemic Merit." *Philosophy of Science* 73 (1): 66–89. <https://doi.org/10.1086/510175>.
- Xie, Feng, and Ting Zhou. 2022. "Industry Sponsorship Bias in Cost Effectiveness Analysis: Registry Based Analysis." *BMJ* 377: e069573. <https://doi.org/10.1136/bmj-2021-069573>.

EPISTEMIC INTIMIDATION AND ILLEGITIMATE VALUE INFLUENCES IN SCIENCE

Manuela Fernández Pinto and Anna Leuschner

27.1 Introduction

In recent years, hostility towards science has been on the rise, a deeply concerning trend for the scientific community itself ([Bluemel 2023](#); [Nogrady 2021](#)). Scientists investigating and stressing the importance and recalcitrance of social, health, and environmental problems, such as anthropogenic climate change, the massive use of pesticides, COVID-19, or the issues of racism or sexism, have found themselves increasingly under attack: Their credibility and integrity have been challenged, they are insulted, ridiculed, and threatened.

According to a recent survey of the members of the American Academy for the Advancement of Science, encompassing a wide range of scientific disciplines, more than half of researchers (51 percent) reported experiencing at least one kind of harassment in their careers, and in many cases, these intimidatory practices had been going on for years, even decades ([O’Grady and Errington 2022](#)). A further study, conducted by an alliance of some of the most important research institutions in Germany, came to similar conclusions: Attacks against scientists have been occurring for some time, but they are currently on the rise ([Bluemel 2023](#)). According to another survey conducted by *Nature*, two-thirds of scientists commenting publicly on COVID research during the pandemic reported negative experiences as a consequence, 22 percent received threats, physical or sexual, 15 percent received death threats, and six reported being attacked ([Nogrady 2021](#)). Similarly, scholars working on gender and race studies have increasingly become fair game for right-wing populists around the world ([Redden 2018](#)).

Such attacks are often framed as mere instances of criticism, invoking freedom of speech and extolling the epistemic value of controversial debate. There is no doubt that the free exchange of opinions and mutual criticism are crucial for the advancement of knowledge. However, justifying these attacks against (specific) scientists under the pretext of free speech is deeply problematic. After all, under this pseudo-idealistic guise, the attacks are clearly directed at silencing dissenting voices and manipulating public discourse to serve particular political or economic agendas. Moreover, the attacks against scientists not only have severe consequences for their personal and professional well-being, but, as we aim to show, they also lead to significant epistemic effects, compromising the production of scientific knowledge, thus causing epistemic harm.

This calls for urgent attention from the philosophy of science, because the frequent intimidation of scientists as an increasing concern for scientific communities is creating an “atmosphere of

intimidation” in science, as climate scientist Michael Mann (2015) once put it, i.e., a social atmosphere that affects the dynamics of scientific practice and, thus, scientific progress. This is connected to the fact that in these cases, political and economic values negatively influence the development of scientific research and the production of scientific knowledge. The issue is connected to the central philosophical debate on science and values, especially when it comes to *the new demarcation problem* (Holman and Wilholt 2022), i.e., the question of how to distinguish between legitimate and illegitimate influences of political and economic values in science today.

Within the science and values debate, Holman and Wilholt identify five strategies for differentiating legitimate from illegitimate value influences in science: *axiological demarcation strategies* that aim to determine what values may legitimately play a role in scientific knowledge processes (e.g., democratically legitimized values); *functionalist demarcation strategies* that focus on the role that values may legitimately play in scientific knowledge processes (e.g., not the role of excluding or defining certain research results in advance); *consequentialist demarcation strategies* that focus on what consequences research decisions may legitimately have (e.g., democratically legitimized consequences); *coordinative demarcation strategies*, according to which value influence is legitimate if it does not mislead the expectations of others (e.g., by pretending that something has been found out, although this is not actually the case); and *systemic demarcation strategies*, according to which illegitimate value influence can only be avoided at a structural level (e.g., through self-regulating on a “marketplace of ideas”). Holman and Wilholt point out that these five strategies are not mutually exclusive; many philosophers of science have argued in favor of more than one of these. In exploring how science can be manipulated and corrupted by intimidatory practices, we seek to shed light on a type of value influence imposed on research by powerful industrial and political stakeholders that is illegitimate according to all five demarcation strategies.

Accordingly, the main purpose of this chapter is to address the uptick of attacks on scientists and highlight their detrimental effects on the production of scientific knowledge. First, we present the contemporary antecedents of the problem, which can be found as early as the 1950s, when the tobacco industry began strategically attacking research showing a link between smoking and lung cancer. Second, we introduce the concept of *epistemic intimidation* to capture the intimidatory practices that specifically target someone (including scientists) in their capacity as epistemic agents. We argue that such epistemic intimidation becomes a genuine threat to science when it is systematically orchestrated by politically and economically powerful groups. Third, we introduce the terminology of epistemic virtues to further sharpen our analysis. We argue that epistemic intimidation leads to the decline of certain fundamental epistemic virtues necessary for the prosperity of science. We then illustrate the problem by examining three ways in which this hinders scientific practice and point out how this is related to value-laden decisions. To conclude, we consider some of the possible strategies that could be implemented for preventing and counteracting such intimidation.

27.2 Contemporary antecedents

The hostility against science that we are experiencing today is not a fundamentally new phenomenon. Historically, scientists who presented politically unwelcome results often did not receive the warmest welcome (Rosenstock and Lee 2002), and even in the current day, attacks initiated by private companies in conjunction with their political allies have a history dating back around 75 years. Indeed, as early as the 1950s, the tobacco industry began strategically targeting research suggesting a link between smoking and lung cancer. The goal was to spread public doubt about the reliability of the science so that tobacco consumers would not be confused or discouraged, and, of course, to obstruct regulatory measures, i.e., either to prevent or at least to delay their proper implementation.

Because the tobacco industry was the first to produce doubt about scientific findings in this way and on this scale, this approach is also known as the “tobacco strategy” (Oreskes and Conway 2010). This is an umbrella term encompassing various tactics used to undermine confidence in science and hinder research, which include overemphasizing uncertainty and promoting bogus controversies. For instance, the tobacco industry created its own institute, the—aptly named—Council of Tobacco Research (CTR), with the mission of producing pseudo-scientific dissent and manufacturing doubt about scientific findings related to the hazards of smoking. Another infamous tactic, which is of the greatest concern to us here, worked more on an *ad hominem* or rather *ad scientificum* basis, consisting of attacking scientists whose research was unfavorable to the industry, for example, by threatening, insulting, or damaging their reputation (e.g., Oreskes and Conway 2010, 137–8).

The tobacco strategy was extremely successful, and it is well documented how the various tactics of the tobacco strategy were then adopted by other industries in the decades that followed. Another early, well-known case is the aggressive campaign against environmentalist Rachel Carson. In her 1962 book, *Silent Spring*, she pointed out the dangerous environmental and health effects of pesticides, particularly Dichlorodiphenyltrichloroethane (DDT). The book contributed significantly to the rise of the environmental movement, meaning conversely that Carson was subjected to aggressive attacks on behalf of the pesticide industry. On the one hand, she was accused of scientific misconduct; on the other, she was berated in newspapers as a “fanatic,” a “hysteric,” an “alarmist,” a “communist,” or simply an “unmarried woman”—always with the aim of discrediting her as a scientist (Bosso 1987; Nieves 2018).

Even though such hostility toward science is nothing new, current intimidatory practices have increased in scope and scale, especially due to the use of online platforms and social media. This has the potential to threaten the viability of science as an epistemic institution in the long run. It has become much easier to spread fake news or to threaten, insult, or offend someone—often under the protection of anonymity. And, populism, which is itself having a comeback in Western democracies, has exacerbated the problem. Thus, attacks against scientists have intensified both quantitatively (due to the sheer number of them facilitated by online platforms) and qualitatively (as they have become more violent).

By far, scientists who endanger the interests of influential economic or political groups with their results are the most common targets of intimidation. In the broadest sense, one can say that scientists are most frequently attacked when their research results shake up existing power relations and point to social change, including regulatory measures in public policy. Such research encompasses fields such as climate science, toxicology and biochemistry, pharmaceutical research, but also certain areas of the social sciences, such as critical race theory or gender studies. This has now been well explored, particularly for environmental and health sciences (e.g., Dunlap and McCright 2010; Malkan 2022; Mann 2015; O’Grady and Errington 2022; Oreskes and Conway 2010; Rohr 2021), but there is also increasing evidence for other fields such as gender studies (e.g., Bluemel 2023).

27.3 Epistemic intimidation

To address the question of how the intimidation of scientists affects the production and dissemination of scientific knowledge, we first wish to suggest that intimidation can be epistemically detrimental in a general sense. This is the case when it targets someone in their capacity as an epistemic agent, i.e., a knower, an inquirer, a reasoner, thus leading to epistemic damage. Here we draw from Fricker’s definition of epistemic injustice (2007, 20). As Fricker further elucidates, the “central cases of epistemic injustice” are based on “tracker prejudices,” which make them “systematic” in that they affect the targeted person throughout all the dimensions of their lives (Fricker 2007, 29). Similarly, we also consider that the “central,” most detrimental cases of epistemic intimidation are “systematic,” but in

a different sense, that is, insofar as they are orchestrated, i.e., methodical, organized, and planned. Moreover, we are especially interested in systematic intimidation linked to structural power relations in society and the maintenance of the status quo, i.e., when systematic epistemic intimidation is the result of organized efforts on behalf of sufficiently powerful political or economic groups who see their predominant views threatened by certain scientific findings, and who respond with orchestrated attacks to defend a status quo that is beneficial to them. Accordingly, we focus on systematic epistemic intimidation, which targets someone in their capacity as an epistemic agent, leads to epistemic harm, and is orchestrated by politically and economically powerful groups.

Conversely, then, there may well be epistemic intimidation that is not systematic, or that does not pertain to efforts of powerful groups. For example, currently in the media, there are those who feel attacked by an alleged “cancel culture.” These may also be cases of epistemic intimidation, but they are not our central concern.¹

At this point, it is fair to ask under what conditions a group of attackers is “sufficiently powerful.” Admittedly, this cannot be precisely defined, making it a potential point of controversy. However, there are certain indicators that need to be scrutinized on a case-by-case basis (such as access to extensive financial resources or political influence via far-reaching connections to social elites) and which may well indicate whether a particular position under attack can in fact be silenced. In cases of systematic epistemic intimidation, the aggressors are part of larger networks that enable them to manipulate the public debate and thus exert considerable influence on economic and political decisions (Kamola 2019; Stanley 2015).

For example, if we look at the now well-examined attacks on climate scientists, we can see how industry executives, certain politicians, and media networks have coordinated attacks on climate scientists, supported by think tanks subsidized by those same industries (Mann 2012; Oreskes and Conway 2010). This then would be an exemplary case of epistemic intimidation.

What is crucial about these cases is that, given existing power relations, they significantly contribute to a deeply hostile atmosphere, where scientists are intimidated as epistemic agents to the extent that it has an overall impact on scientific practice within their research fields. But how exactly does this work? In the following, we will explore the question of how these attacks affect the production and dissemination of scientific knowledge. To do so, we draw on virtue epistemology and the debate over inductive risk before illustrating the problem by way of some examples.

27.4 Epistemic virtues and the problem of inductive risk

While traditional epistemology focuses primarily on the analysis of key concepts such as “knowledge,” “justification,” or “truth,” virtue epistemology places intellectual actors and communities at the center of epistemic evaluation. Particular attention is paid to the epistemic virtues (or vices) embodied by these actors. Epistemic virtues and vices can mean quite different character traits and dispositions, depending on the context. For instance, epistemic virtues include traits such as creativity and curiosity, as well as intellectual honesty and modesty. Epistemic vices, in turn, would include intellectual arrogance, cowardice, narrow-mindedness, or naivety.²

Obviously, certain epistemic virtues—such as intellectual courage, open-mindedness, integrity, disinterest, honesty, impartiality, intellectual sobriety, creativity, or imagination—are particularly important in the context of scientific knowledge acquisition (Paternotte and Ivanova 2017). Curiosity, intellectual resilience, and perseverance are further epistemic virtues that appear to be crucial in scientific contexts.

The conceptual framework of virtue epistemology is helpful for understanding how attacks on scientists affect their work. It seems obvious that all of these epistemic virtues are impacted by strategies of intimidation. An intimidated scientist can find it harder to show epistemic courage and

open-mindedness. Similarly, integrity, altruism, honesty, and impartiality are hindered when intimidation prevails over the free exchange of ideas. Creativity, imagination, and curiosity also suffer. Similarly, intellectual resilience and persistence are impaired, as the insistence on findings could provoke further attacks. And even intellectual sobriety is affected, because intimidation can cloud one's perspective and, thus, often goes hand in hand with biased assessments of research subjects. In sum, the epistemic virtues that are particularly important for scientific work obviously are all undermined by epistemic intimidation.³

This can significantly impact the conduct of research and its outcomes. The central consequence is that scientists engage in self-censorship, for example, by avoiding certain research questions (especially young scientists can be discouraged from pursuing certain lines of research), or by hesitating to defend certain hypotheses with the vigor they consider appropriate in light of the available empirical evidence, which in the worst case can lead to biased research results from a continued preference for certain risks.

As [Hempel \(1965\)](#) emphasized, in empirical research, a gap always remains between the probability of a hypothesis that can be inferred from the available empirical evidence and the certainty needed to accept the hypothesis. Even in the best case, some risk of error always remains, and this risk is what Hempel calls the inductive risk. The inductive risk, then, is the risk of either accepting a hypothesis even though it is actually false (false-positive error) or rejecting the hypothesis even though it is actually true (false-negative error). Of course, scientists seek to avoid both errors as much as possible, but this necessarily involves weighing the risk of making a false-positive against the risk of making a false-negative error. Unavoidably, the scientists are required to make a value judgment ([Douglas 2009](#); [Rudner 1953](#)).

Now, it should be noted that there is a general tendency in science to favor false-negative errors simply because scientists wish to avoid asserting correlations where there are none and then having to retract their assertion completely. After all, it is much more embarrassing to have to retract a finding later (i.e., to have made a false-positive error) than to have to admit that one may have been too reticent (i.e., to have made a false-negative error). This tendency also has a name: *erring on the side of least drama* ([Brysse et al. 2013](#)). Notice that the decision to avoid false-positive errors over false-negative errors reflects a tendency to serve some social values over others (i.e., to prefer public risks over producer risks, e.g., not being as protective of community members who could be exposed to hazards and being more protective of “producers” who could be engaging in hazardous activities).

Focusing on climate research, [Brysse et al. \(2013\)](#) have already indicated that the tendency to favor false-negative errors is exacerbated by attacks on climate scientists because of the growing fear of a possible false-positive error: Since climate scientists are accused of being alarmist, hysterical, and of exaggerating, they become particularly cautious because—in line with the principle of stereotype threat—they do not want to confirm this accusation under any circumstances (see also [Lewandowsky et al. 2015](#)).

Analogously, in cases of epistemic intimidation, aggressive lobbying groups assume that scientists are ringing alarm bells regarding problems or dangers that do not exist (or at least do not exist in that form), such as problems of sexism and racism, or dangers from the use of tobacco, pesticides, or fossil fuels. The insinuation that the problems or dangers do not even exist is accompanied by attacks (spreading public doubts about scientific competence and integrity, ridiculing, and issuing personal threats), and the reaction to these attacks, which is increasingly evident in interviews and surveys, is, unsurprisingly, that concerned scientists tend to become overly cautious and reluctant in how they conduct their research. We will now show in more detail the negative impact of epistemic intimidation on the production and communication of scientific knowledge.

27.5 The effects of epistemic intimidation

Epistemic intimidation can have a devastating effect on science. We will focus here on what we consider the three salient ways in which it can impact scientific practice: (1) by encouraging scientists to become overly cautious, (2) by creating a “chilling effect,” in which scientists become reluctant to conduct research on certain topics or disseminate specific results, and (3) by contributing to the loss of competent researchers.

First, let us consider the point made by [Brysse et al. \(2013\)](#) again. As they noted, there is a tendency for scientists to *err on the side of least drama*. Epistemic intimidation, however, can push this tendency too far, leading scientists to a “conservative bias,” in this case, to systematically underestimate the impacts of climate change. To show this, [Brysse et al. \(2013\)](#) begin by examining a series of studies contrasting the empirical evidence with what reports by the Intergovernmental Panel on Climate Change (IPCC) have claimed over the years, including projections of global mean temperature change, sea level rise, and CO₂ concentration in the atmosphere, as well as rainfall increase, arctic sea-ice decline, melting of the continental ice-sheet, and hurricane intensity and frequency ([Allison, Bindoff, and Bindschadler 2011](#); [Hansen 2007](#); [Rahmstorf, Cazenave, and Church 2007](#)). As a result of their survey, they conclude that “in many cases IPCC predictions seem to have underestimated actual outcomes” ([Brysse et al. 2013](#), 329). Overall, they suggest that the constant underestimation of climate change impacts can be due at least in part to the hostile attacks against climate science, and they reconfirm this in a more recent publication ([Oppenheimer, Oreskes, and Jamieson 2019](#)). [Leuschner \(2018\)](#) finds a similar trend when examining the history of the Burning Embers Diagram, which is used to illustrate five central risk groups in the context of climate change in IPCC reports since 2001. She argues that “there has been an apparent tendency to systematically neglect negative findings in the IPCC reports” (2018, 1260).

This tendency toward overcaution is also supported by testimonies from scientists themselves. For example, climate scientists have stated that in order to avoid attacks they are “reluctant to make strong claims about the scientific evidence,” that they prefer to “err on the side of conservatism,” (Oreskes and Conway 2010, 264–5) and that they “keep a low profile and go with the flow” ([Bradley 2011](#), 137). Similarly, a survey of climate scientists in New Zealand and the UK concluded that 60 percent of scientists surveyed had become overly cautious as a result of attacks ([Sharman 2015](#)). As indicated by a current survey on scientists in Germany from a broad range of disciplines, this effect can also be found in other disciplines ([Bluemel 2023](#)).

This is problematic for science, of course, but being overly cautious (as opposed to being just cautious), can also have dramatic consequences for the public. In the case of climate science, for instance, the consistent underestimation of climate change impacts can contribute to delaying timely climate change mitigation measures, which in turn can have devastating social and environmental impacts. In other words, scientists end up adopting the value preferences of the industry in that they prioritize some social values, favoring the fossil fuel industry (e.g., not spending money on potentially unnecessary mitigation or adaptation measures) over other social values (e.g., protecting the public from potentially serious climate) and assess inductive risks along these lines.

However, overcaution is just one of many possible effects of systematic intimidation in science. Overall, intimidation can lead to a “chilling effect” in scientific practice (our second point mentioned above). [Kempner \(2008\)](#) introduced the idea of a “chilling effect” to capture practices of self-censorship that he uncovered among scientists working on politically controversial topics. This pertained particularly to the rescinding of National Institutes of Health (NIH) funding for the study of sexual behavior due to pressure from Republican members of the US Congress and further intimidation from the conservative Christian group Traditional Values Coalition. The study uncovered a series of self-censorship strategies ranging from complete silence to minor

modifications and even omissions. Most scientists were found to disguise the most controversial aspects of their research in grant applications: 51 percent, for instance, avoided using certain “red flag” words, such as “gay,” “homosexual,” and “AIDS” (Kempner 2008, 1575). More significantly, for at least 24 percent of researchers, self-censoring went beyond language tuning, for example, avoiding research on marginalized populations altogether, and dropping or not renewing projects considered politically controversial, such as abortion, emergency contraception, and condom use (Kempner 2008, 1575). Four out of 82 researchers reported making radical changes to their careers:

Two left research positions, in which they had to raise their own salaries by securing grant money for the security of research jobs with guaranteed salaries. A third continued their research, but ‘left the country for a more supportive science environment’ (S). The fourth left academic research altogether, declaring that ‘This [controversy] ended my research career.’
(Kempner 2008, 1576)

Another survey by Pieter-Paul Verhaeghe (2022) came to similar results: Researchers were asked to what extent intimidation affected (1) their communication of research results, (2) the study of particular topics, (3) voicing their opinion on particular topics, as well as (4) their decision to stay in academia, and (5) to teach certain topics. A total of 38 percent showed hesitancy toward voicing their opinions on particular topics, and around 28 percent felt both reluctant to communicate their research findings and to stay in academia as a result (2023, 28). Moreover, they found significant and strong associations between having suffered intimidation and agreeing with these five consequences, especially hesitancy toward staying in academia. Verhaeghe concludes: “The more one is intimidated or harassed as a scientist, the more far-reaching the academic consequences” (2022, 29). Further studies confirm the relation between being a target of intimidation and the chilling effect among academics (Sharman 2015; Väliverronen and Saikkonen 2021). As a result of intimidation, scientists end up inadvertently favoring the political or cultural values of those who don’t want a particular topic to be studied or discussed, which results in a skewing of the topics being studied.

As for the third point in our list, all these studies have shown that epistemic intimidation leads to a loss of talented individuals who would rather leave their careers in science than further burden themselves with these attacks and the hostile climate that ensues. As Kempner (2008) argues, leaving academia is one of the strongest effects and consequently rather rare. In his sample of 82 PIs, only one person made this decision. However, although this does not appear to be many, we would note that there could be a sample bias in surveys exploring the effects of hostility against scientists, since those who leave academia might be neglected in the surveys. If this is true, the real number of those leaving academia, or of students deciding not to start an academic career at all, due to (anticipated) hostility, would be higher than the surveys indicate. This is consistent with Verhaeghe’s (2022) findings, in which more than a quarter of researchers reported considering leaving academia as a result of intimidation.

If we take into account the vast personal and social effort required to train someone as a scientific researcher, as well as the extent to which contemporary societies depend on scientific expertise and innovation, then losing talent has to be understood as one of the most glaring epistemic consequences of the systematic intimidation of scientists. This is similar to the “brain drain” issue, i.e., the problem of national scientific talent leaving low- and middle-income countries for better opportunities in richer countries with more promising research environments, due to, among other factors, political instability (Docquier, Lohest, and Marfouk 2007). Perhaps we are not there yet, but if it becomes a tendency for researchers to leave academia for places with more promising and friendly working environments, instead of resisting intimidation and

enacting change, this could lead to the reinforcement and normalization of epistemic intimidation, which in turn would compromise the proper functioning of science as society's paramount epistemic institution.

Considering the significant implications, it is imperative for a society to actively nurture and support its talent pool to the fullest extent possible. To ensure the continued flourishing of science, societies must foster a safe research environment that cultivates epistemic virtues, so that unbalanced value preferences, and thus biases, do not emerge. This involves creating a climate where researchers feel supported in their epistemic endeavors.

In sum, epistemic intimidation has harmful effects on science, as the examples of overcaution, the chilling effect, and the loss of talent suggest: It leads to value influence that is illegitimate according to all five "demarcation strategies" identified by [Holman and Wilholt \(2022\)](#), since it results in choices of methods and goals that are obviously not in line with democratic values, skews intra-scientific processes in various ways, and undermines both coordinative societal processes and scientific self-regulation. Thus, it is not only the moral and social well-being of individual scientists that is compromised (as awful as this can be), but the thriving and integrity of the scientific process itself.

27.6 Conclusion

In conclusion, epistemic intimidation represents a critical challenge to the integrity and progress of scientific inquiry. As evidenced by the increased prevalence of harassment and attacks targeting researchers in climate science, but also across various other disciplines, the repercussions extend far beyond individual well-being, undermining the very foundation of knowledge production.

The notion of systematic intimidation underscores the deliberate nature of these tactics, designed to silence dissenting voices, and manipulates public discourse for political or economic gain. This manipulation not only stifles scientific advancement but also distorts the relationship between science and society, eroding trust and inhibiting the pursuit of knowledge.

Within the framework of virtue epistemology, it becomes clear that these intimidation practices foster a climate of fear among scientists, hindering the cultivation of epistemic virtues that are fundamental to science. The consequences are particularly stark in fields required for policy advice such as "regulatory science" or climate science. In the latter case, it is already well documented that the persistent intimidation of researchers has indeed contributed to the downplaying of urgent issues: It has led to the negligence of certain industry-unfriendly research topics and questions as well as to an exaggerated preference for false-negatives over false-positives and, thus, to a lopsided distribution of inductive risk. Taking all this into account, and insofar as these attacks are orchestrated by groups motivated by political and economic interests, we have come to the conclusion that epistemic intimidation promotes illegitimate value influence on scientific research according to all five demarcation strategies identified by Holman and Wilholt.

Finally, we wish to present some promising suggestions on how this problem could be countered. Addressing epistemic intimidation demands a multifaceted approach, involving not only individual engagement but also institutional and societal interventions. Strategies to prevent and counteract intimidation may include implementing and bolstering support networks for targeted individuals, promoting transparency and accountability in scientific communication, and advocating for policies that safeguard academic freedom and integrity. Targeted scientists agree that institutional support is a deciding factor in determining the extent of the epistemic damage caused by intimidatory practices. However, cases often remain unreported because targeted researchers are self-blaming, or because they feel ashamed and are concerned about damaging their reputation ([Branford, Grahle, and Heilinger 2019](#); [Veletsianos et al. 2018](#)). Therefore, many academic institutions are oblivious to the intimidation suffered by their faculty members ([Cassidy, Faucher, and Jackson 2017](#)).

This situation could potentially improve if academic institutions offered support. Universities especially could develop and promote strategies to support their researchers when they are under attack. Such commitments could help to reduce concerns that one may be targeted when conducting certain research, and they could also help to encourage researchers to report cases of attacks. Some of the more promising strategies that have been suggested are as follows:

- 1 **Policies:** Establish and enforce clear policies and procedures against the intimidation of faculty members, including a standardized response to communicate quickly when a faculty member is being targeted. Ideally, this response would assert the expertise and value of the scholar for the institution, endorse academic freedom as fundamental, acknowledge the importance of diversity of views and academic debate, while unequivocally condemning intimidation (AAUP 2022).
- 2 **Training:** Offer training sessions for faculty and staff on identifying and dealing with intimidation, including how to report incidents and access support resources. Another key preventive strategy is offering social media training for faculty to protect themselves from cyberbullying, including two-factor authentication and password management apps (Veletsianos et al. 2018).
- 3 **Support Networks:** Help establish support networks or committees for targeted faculty members by providing confidential assistance, connecting them to unions, peer support, or counseling services, helping them monitor messages so they need not do it themselves, and removing contact information from the institutional website and directory, among others (AAUP 2022).
- 4 **Legal Resources:** Provide access to legal assistance and information on legal protections, such as academic freedom, harassment law, and copyright law, and also to information regarding public records requests that apply to the institution (AAUP 2022).

By implementing such strategies, higher education institutions could create safer and more supportive environments for faculty members, fostering academic freedom, intellectual inquiry, and the pursuit of knowledge. However, further insights into the tactics of epistemic intimidation and its effects are needed, as well as further debate and research on effective ways of counteracting it. Moreover, we deem it an important task for philosophy to promote discussions on how to counteract the spreading of an “atmosphere of intimidation,” in both science and society, and thus to reinforce democracy and to further explore how to improve social conditions for science.

Acknowledgment

This article is the outcome of a joint research project of the authors, funded by the German Research Foundation (DFG).

Notes

- 1 The term “cancel culture” is used to claim that conservative and right-wing public figures, including academics, are victims of silencing by groups that are in fact structurally powerless, such as trans people or people of color. Without doubt, attacks from minority groups can also lead to the intimidation of individuals; in some cases, they might even have epistemic effects beyond the individual level. However, given predominant power relations and wide-spread biases in society overall, anti-feminist, trans-critical, or climate-skeptic positions, for example, find great financial and ideological support. Thus, we take it that, given the current political situation in most western democracies, the most detrimental cases of epistemic intimidation tend to originate from actors and movements advocating these positions.
- 2 There is a growing body of literature on epistemic virtues and vices, including virtues and vices not only at the individual, but also on the social level (Kidd et al. 2023; Turri et al. 2021). However, for the purpose of this chapter, it is sufficient to focus on exemplary individual epistemic virtues. In addition, as Elliott and Korf

(2024) suggest, insofar as epistemic virtues can be understood as something that we desire or prioritize, they would also count as “values.” In this sense, there is a substantial link between virtue epistemology as applied to science and the literature on science and values.

- 3 One might expect that epistemic intimidation even leads to scientists within attacked communities being pushed to adopt epistemic vices. For now, we focus here on how epistemic intimidation affects the acquisition and exercise of epistemic virtues that are crucial for scientific research.

References

- AAUP. 2022. “Faculty First Responders Webinar.” Video: <https://www.youtube.com/watch?v=JaHINlfXvJw>.
- Allison, Ian, Nathan L. Bindoff, Robert A. Bindschadler, et al. 2011. *The Copenhagen Diagnosis: Updating the World on the Latest Climate Science*. Elsevier.
- Blumel, Clemens. 2023. *Anfeindungen gegen Forschende*. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung. https://scicomm-support.de/wp-content/uploads/2024/05/2024_05_16_Studie-Anfeindungen-gegen-Forschende_KAPAZ.pdf.
- Bosso, Christopher. 1987. *Pesticides and Politics: The Life Cycle of a Public Issue*. University of Pittsburgh Press.
- Bradley, Raymond. 2011. *Global Warming and Political Intimidation: How Politicians Cracked Down on Scientists as the Earth Heated Up*. University of Massachusetts Press.
- Branford, Jason, André Grahle, Jan-Christoph Heilinger, et al. 2019. “Cyberhate against Academics.” In *Responsibility for Refugee and Migrant Integration*, edited by Karly Kehoe, Eva Alisic, and Jan-Christoph Heilinger. De Gruyter.
- Brysse, Keynyn, Naomi Oreskes, Jessica O’Reilly, and Michael Oppenheimer. 2013. “Climate Change Prediction: Erring on the Side of Least Drama?” *Global Environmental Change* 23 (1): 327–37. <https://doi.org/10.1016/j.gloenvcha.2012.10.008>.
- Cassidy, Wanda, Chantal Faucher, and Margaret Jackson. 2017. “Adversity in University: Cyberbullying and Its Impacts on Students, Faculty and Administrators.” *International Journal of Environmental Research and Public Health* 14 (8): 888. <https://doi.org/10.3390/ijerph14080888>.
- Docquier, Frédéric, Olivier Lohest, and Abdeslam Marfouk. 2007. “Brain Drain in Developing Countries.” *The World Bank Economic Review* 21 (2): 193–218. <https://doi.org/10.1093/wber/lhm008>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Dunlap, Riley E., and Aaron M. McCright. 2010. “Climate Change Denial: Sources, Actors and Strategies.” In *Routledge Handbook of Climate Change and Society*, edited by Constance Lever-Tracy. Routledge.
- Elliott, Kevin C., and Rebecca Korf. 2024. “Values in Science: What Are Values, Anyways?” *European Journal for Philosophy of Science* 14. <https://doi.org/10.1007/s13194-024-00615-3>.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
- Hansen, James. 2007. “Scientific Reticence and Sea Level Rise.” *Environmental Research Letters* 2: 024002. <https://doi.org/10.1088/1748-9326/2/2/024002>.
- Hempel, Carl G. 1965. “Science and Human Values.” In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. The Free Press.
- Holman, Bennett, and Torsten Wilholt. 2022. “The New Demarcation Problem.” *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Kamola, Isaac. 2019. “Dear Administrators: To Protect Your Faculty from Right-Wing Attacks, Follow the Money.” *Journal of Academic Freedom* 10: 1–24.
- Kempner, Joanna. 2008. “The Chilling Effect: How Do Researchers React to Controversy?” *PLOS Medicine* 5 (11): 1571–78.
- Kidd, Ian James, Heather D. Battaly, and Quassim Cassam, eds. 2023. *Vice Epistemology*. Routledge.
- Leuschner, Anna. 2018. “Is It Appropriate to ‘Target’ Inappropriate Dissent? On the Normative Consequences of Climate Skepticism.” *Synthese* 195 (3): 1255–71. <https://doi.org/10.1007/s11229-016-1267-x>.
- Lewandowsky, Stephan, Naomi Oreskes, James S. Risbey, Ben R. Newell, and Michael Smithson. 2015. “Seepage: Climate Change Denial and Its Effect on the Scientific Community.” *Global Environmental Change* 33: 1–13. <https://doi.org/10.1016/j.gloenvcha.2015.02.013>.
- Malkan, Stacy. 2022. *Merchants of Poison: How Monsanto Sold the World on a Toxic Pesticide*. https://oregontruffletryst.com/Merchants_of_Poison_Report_final_120522.pdf.
- Mann, Michael. 2012. *The Hockey Stick and the Climate Wars: Dispatches from the Front Lines*. Columbia University Press.

- Mann, Michael. 2015. "The Serengeti Strategy: How Special Interests Try to Intimidate Scientists, and How Best to Fight Back." *Bulletin of the Atomic Scientists* 71 (1): 33–45. <https://doi.org/10.1177/0096340214563674>.
- Nieves, Autumn. 2018. "Rachel Carson: Proving the Competency of Femininity." *Women Leading Change: Case Studies on Women, Gender, and Feminism* 3 (2): 52–64.
- Nogrady, Bianca. 2021. "'I Hope You Die': How the COVID Pandemic Unleashed Attacks on Scientists." *Nature*, October 13. <https://www.nature.com/articles/d41586-021-02741-x>.
- O'Grady, Cathleen, and Tim Errington. 2022. "'In the Line of Fire': Science News Survey Results (AAAS Members)." *Science*, March 25. https://www.science.org/pb-assets/science.abq1755/aaas_analysis-1650295375413.html.
- Oppenheimer, Michael, Naomi Oreskes, Dale Jamieson, et al. 2019. *Discerning Experts: The Practices of Scientific Assessment for Environmental Policy*. The University of Chicago Press.
- Oreskes, Naomi, and Erik Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury Press.
- Paternotte, Cedric, and Milena Ivanova. 2017. "Virtues and Vices in Scientific Practice." *Synthese* 194 (5): 1787–807. <https://doi.org/10.1007/s11229-016-1023-2>.
- Rahmstorf, Stefan, Anny Cazenave, John A. Church, et al. 2007. "Recent Climate Observations Compared to Projections." *Science* 316 (5825): 709. <https://doi.org/10.1126/science.1136843>.
- Redden, Elizabeth. 2018. "Global Attack on Gender Studies." *Inside Higher Ed*, December 5. <https://www.insidehighered.com/news/2018/12/05/gender-studies-scholars-say-field-coming-under-attack-many-countries-around-globe>.
- Rohr, Jason R. 2021. "The Atrazine Saga and Its Importance to the Future of Toxicology, Science, and Environmental and Human Health." *Environmental Toxicology and Chemistry* 40 (6): 1544–58. <https://doi.org/10.1002/etc.5037>.
- Rosenstock, Linda, and Lore Jackson Lee. 2002. "Attacks on Science: The Risks to Evidence-Based Policy." *American Journal of Public Health* 92 (1): 14–18. <https://doi.org/10.2105/AJPH.92.1.14>.
- Rudner, Richard. 1953. "The Scientist *Qua* Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Sharman, Amelia. 2015. "The Impact of Controversy on The Production of Scientific Knowledge." *Centre for Climate Change Economics and Policy*. Working Paper 233.
- Stanley, Jason. 2015. *How Propaganda Works*. Princeton University Press.
- Turri, John, Mark Alfano, and John Greco. 2021. "Virtue Epistemology." In *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2021/entries/epistemology-virtue/>.
- Väliverronen, Esa, and Sampsa Saikkonen. 2021. "Science Communicators Intimidated: Researchers' Freedom of Expression and The Rise of Authoritarian Populism." *Journal of Science Communication* 20 (4): 1–18. <https://doi.org/10.22323/2.20040208>.
- Veletsianos, George, Shandell Houlden, Jaigris Hodson, and Chandell Gosse. 2018. "Women Scholars' Experiences with Online Harassment and Abuse: Self-Protection, Resistance, Acceptance, and Self-Blame." *New Media & Society* 20 (12): 4689–708. <https://doi.org/10.1177/1461444818781324>.
- Verhaeghe, Pieter-Paul. 2022. *The Academic Intimidation and Harassment of Scientists at Flemish Universities in Belgium*. VUB.

28

NETWORKS AND VALUES

Rebecca Korf and Cailin O'Connor

28.1 Introduction

The values in science literature has considered deep questions around value-laden aspects of inquiry like the following: Where and when can values of different sorts legitimately influence science? What sorts of values, and related ethical practices, best promote scientific inquiry? How do individual beliefs and biases impact science? Earlier work focused mostly on how values interact with individual practice. But increasingly there has been a move toward thinking about how the impacts of values can reverberate through scientific communities and how the dynamics of social groups may lead to unexpected outcomes when values shape science. At the same time, social epistemologists and others have been using network models to think about how social network structures and group features shape knowledge generation. These models are often useful in thinking about the emergent, group-level effects of values on scientific progress.

This chapter will briefly survey this network literature and connect it with the literature on values in science. As we will see, the causal links between values and epistemic outcomes are not always simple or predictable. Seemingly unproblematic types of value influence can sometimes have unexpected negative impacts on epistemic outcomes, and seemingly problematic sorts of value influence can sometimes improve outcomes. In other cases, network models reveal that policy proposals to improve scientific functioning may not work as suggested. Still other models point to places where new work on values in science may be useful.

A central take-away is that the impacts of value influence are often dependent on emergent, group-level dynamics, which are hard to predict. Thus, philosophers should take these sorts of phenomena into account, and generally take care, when thinking about value influence. This sort of caution is especially important in the domain of policy discussions where philosophical reasoning may impact real-world decision-making.

The rest of the chapter will proceed as follows. [Section 28.2](#) briefly surveys relevant literature on values in science. The heart of this chapter is [Section 28.3](#) which discusses results from network models and their relevance to the values in science literature. And [Section 28.4](#) briefly concludes.

28.2 Values in science

The values and science literature is concerned with both epistemic impacts and ethical issues arising from value influence in science. While the literature has persuasively argued that science has responsibilities beyond simply providing as much knowledge as possible, it is also driven by the fact that an

important goal of science is to provide helpful knowledge on a desirable time frame (and that there is an ethical responsibility to do so). A major question is how different instances of value influence might impact science's ability to fulfill this goal.

Early work in values and science generally focused on the ways that values or social beliefs held by an individual scientist might affect their decisions at various points of the scientific practice. The classic paper from [Rudner \(1953\)](#), for example, is about value judgments individual scientists make and much work that followed retained this individual lens. Recent prominent proposals for value management strategies like those from [Douglas \(2009\)](#), [Elliott \(2017\)](#), [Steel \(2017\)](#), and [Brown \(2020\)](#) also take the individual scientist as their target.

This individual lens has yielded important insights but, of course, scientists do not work alone. Scientists collaborate with colleagues, interact within communities of researchers, contend with social structures in their disciplines and institutions, work within societies that include political and industrial influences, and are part of wider publics which are interested in and affected by scientific results. For this reason, some researchers have focused more on the group or community level when thinking about value influence in science. Notable early examples of this sort include [Longino \(1990\)](#) and [Solomon \(2001\)](#) who call for objectivity and knowledge to be reframed as social concepts.

Recently philosophers have worked on the impact of institutional structures on values in science. [Douglas \(2018, 2021\)](#), for example, explores which norms are useful for institutions, as well as individuals, in promoting ethical responsibility in science. She also focuses on how to hold scientists and institutions accountable when they fail to act ethically. [Rolin \(2015\)](#) argues that scientific collaboration illuminates how values can act differently on the individual and scientific level.

It is clear that including a social focus is important for the values in science literature. One persistent challenge for socially focused work is that predicting and understanding the dynamics of social groups is notoriously difficult. In the next section we turn to network models, which are specifically designed to help theorists get a handle on phenomena in epistemic communities, like groups of scientists. These models have proved useful to social epistemology—a sub-field with ongoing links to values in science. As we will show, they are useful to values in science as well.

28.3 Values and network models

Social group dynamics are complex and sometimes difficult to predict. In many cases individual decisions, preferences, or traits can lead to surprising emergent group effects. Simple models in the social sciences have been used to demonstrate and underpin this fact. Models are especially helpful here because they are easy to control and understand. Thus, unlike in most complex, real-world communities, researchers using models can definitively show how particular features lead to emergent group-level effects, at least in principle.

The values in science literature is interested in epistemic communities, and especially scientific ones. Work on this sort of community provides ample evidence for emergent, surprising, group-level phenomena, as we will see ([Mayo-Wilson, Zollman, and Danks 2011](#)). Some models demonstrate how value influence on the small scale can add up to emergent group effects. Others are relevant to policies intended to manage value influence. Still others point to under-explored topics where the values in science literature has an important role to play. Here we introduce this literature and discuss its relevance for values in science.

28.3.1 Industrial selection

One important question for values in science is how and whether industrial values—those related to the desire to develop and sell new technologies or products—can legitimately influence science.

Some argue we should use the roles values play in science, and the context in which their influence occurs, to determine the legitimacy of value influence, rather than the source or identity of such values (most prominently, [Douglas \(2009\)](#)). Under this sort of account, the “external” stages of science, including funding choices, can be legitimately influenced by a wide variety of values, as long as “internal” parts of research are not affected. So, for example, a pharmaceutical company could legitimately fund or even directly produce research on the efficacy of a drug, as long as their interests did not directly influence the research process.

As [Holman and Bruner \(2017\)](#) and [Korf \(2025\)](#) argue, though, recent results in network modeling raise worries for this sort of account. [Holman and Bruner \(2017\)](#) develop a network model to demonstrate how industrial influence over research funding can lead to surprising negative outcomes. The type of model they use was developed in economics by [Bala and Goyal \(1998\)](#) and brought into philosophy by [Zollman \(2007\)](#). Agents learn which of two actions, *A* or *B*, is more successful by gathering data about their preferred action and sharing it with network neighbors. Over time, they update their beliefs about the actions in light of data. The agents can represent a community of scientists testing two action-guiding theories and sharing their research over time—such as medical researchers learning whether bacteria or stomach acid is the primary cause of ulcers, whether smoking is likely to cause lung cancer or not, or which of two pills is more efficacious ([Weatherall, O’Connor, and Bruner 2020](#); [Zollman 2007, 2010](#)). We do not go into details of the model for space reasons.

Inspired by historical cases [Holman and Bruner \(2017\)](#) develop a version where agents use different methods to explore the world. These methods are better or worse at tracking the actual underlying quality of the two actions. For instance, one medical researcher might use a highly skewed sample population and thus develop a poor understanding of a drug, while another researcher using a representative sample might get better data. Thus, agents testing *A* and *B* actually draw data from slightly different distributions.

The key feature of their model is that industry agents choose which researchers to fund and, in doing so, shape their eventual influence on the community. Agents with more funding gather more data, which then influences the beliefs of network neighbors. The model also assumes that new agents join the network, and when they do, they tend to adopt the methods of those who have been more productive (i.e., those with more funding).

They consider cases where industry has a stake in what theories scientists adopt. Suppose traditional treatment *B* is actually better than new drug *A*, but industry wants to sell new drug *A*. Industrial actors find and fund only those researchers who happen to use poor methods, i.e., ones that tend to show *A* is better than *B*. While the researchers maintain their academic integrity—they do not alter their methods in light of industry funding—this funding can nonetheless sway the community away from the correct belief.¹

[Holman and Bruner \(2017\)](#) argue their model raises worries for proposals from [Douglas et al. \(2014\)](#) about mitigating industry harms. While the latter authors are mostly focused on how individuals can preserve their integrity in the face of industry influence, personal integrity cannot mitigate industry harms based in selection.

[Korf \(2025\)](#) points out that this case demonstrates a larger feature of science, which raises worries for accounts like that of [Douglas \(2009\)](#). (This account, remember, limits industrial influence to the external parts of science.) When scientists update their beliefs in these models on the results of network neighbors, the relative productivity of different agents impacts how their colleagues later evaluate hypotheses. By changing what a community of scientists believes about the world, industrial selection should change how these scientists understand and evaluate evidence later. Thus, external influence that systematically affects productivity—which grants are funded, who is given prestigious university positions, and which work is chosen to be published, along with industrial funding—ends up shaping internal scientific processes at later stages. Value influence can propagate and reverberate

throughout a community in this way. Hence, accounts that limit certain kinds of value influence to external stages of science may fail to actually exclude such influence on the internal level in the longer term.

In thinking about real-world communities, we might also expect external influence to become internal via other channels. Industrial selection shapes which theories scientists hold by shaping the distribution of research they see. Background beliefs deeply shape research choices at every stage of the scientific process (Anderson 2020; Longino 1990; Okruhlik 1994; Solomon 2001). Without further work we do not know what form these dynamics might take, but expect them to be complex and unpredictable. It will thus be hard to assess up front how external value influence should impact internal decision-making later on.

The models described here, and their surprising, emergent social effects, have several lessons for values in science. (1) Attempts to limit value influence to external stages of science may not prevent unwanted harms. (2) More generally, the demarcation between internal and external influence is not as simple as some have previously thought. And (3) attempts to mitigate industry harms on science cannot focus on individual researcher integrity alone.

28.3.2 Dissent

A topic that has received substantial attention in the values in science literature, as well as in metascience, is dissent (de Melo-Martín and Intemann 2018; Intemann and de Melo-Martín 2014; Paetkau 2024; Solomon 2001). Many worry about harmful dissent by industrial interests, or “epistemically detrimental dissent.” But normal dissent is a crucial part of the scientific process. This leads to two key questions: Can we delineate inappropriate dissent from appropriate dissent? And if so, should we attempt to suppress inappropriate dissent?

Coates (2024) uses a network model to address the latter question, and especially the debate between those who do think it appropriate to suppress cynical dissent (Biddle, Kidd, and Leuschner 2017; Biddle and Leuschner 2015; Cook 2017; Leuschner 2018; Nash 2018; Oreskes 2017) and those who worry that this sort of policy risks suppressing useful dissent in science (de Melo-Martín and Intemann 2014, 2018).

He starts with a network of agents engaging in a two-armed bandit problem. Following Zollman (2009), he assumes that instead of informally sharing results through their social network, scientists read a “journal,” which chooses results to share. He compares the journal (1) randomly selecting results to share and (2) “hiding dissent.” In the latter, results are reviewed by community members. If all the reviewers believe the results support the incorrect theory (on their own lights), they are not shared. Thus, less popular theories tend to be suppressed.

Coates (2024) finds that hiding dissent tends to harm group outcomes. Communities in this sort of model benefit from a period of diversity of practice, where multiple agents test multiple theories long enough to ensure they pick the best one. Hiding dissent tends to decrease diversity of practice. Thus, the community more often preemptively settles on the worse theory. The exception, on his findings, is that after some period of exploration, where the community has already tested both theories extensively, suppressing dissent speeds consensus and does not harm outcomes.

Coates also considers versions of the model where one agent represents some cynical interest, such as an industry representative, following Holman and Bruner (2015). This agent persistently shares data that tends to support the worse theory. Even in this case, Coates finds that hiding dissent is typically harmful as it can quickly push the entire community to the incorrect (industry-supported) consensus.

Notably, though, his model does not include any way for agents to detect epistemically detrimental dissent. Thus, the failure mode in hiding dissent—where the community is influenced toward the

worse belief by industry and starts to suppress the better one—might be avoidable in more realistic cases. [Holman and Bruner \(2015\)](#), in an earlier model, study the impacts of an industrial agent who shares misleading data. They consider situations where other agents can stop listening to the industrial agent by disconnecting from others whose data does not match their beliefs. The result is that the community often learns to ignore the biased agent and performs better.

One take-away might be that even simple learners can successfully detect and avoid some sorts of misleading data. It might very well be possible, with more sophisticated cognitive abilities, to detect and avoid cynical dissenting evidence. In addition, a number of network models have demonstrated the extensive harms of industrial propaganda, lending weight to arguments that industry harms should be taken seriously ([Holman and Bruner 2015, 2017](#); [Lewandowsky et al. 2019](#); [Weatherall, O'Connor, and Bruner 2020](#)). If so, minor risks of suppressing appropriate dissent may be worth taking.

Here models from both [Coates \(2024\)](#) and [Holman and Bruner \(2015\)](#) directly inform debates in values in science over policy proposals by illuminating complex social processes. These models should not be taken to directly predict outcomes in real communities. But they can be used as aids to normal reasoning or rigorous versions of thought experiments, which help inform the effects of policy choices in science.

28.3.3 Open science

The open science movement aims to improve research integrity and access by promoting the open sharing of products of science.² To this point it has generally been assumed that increased sharing, including during intermediate stages of science, will improve research outcomes. Some extant results in network modeling, though, raise worries about unexpected downsides. These results do not directly bear on previous work in values and science on sharing or transparency. Still, we include them here given the general focus in values in science on science communication, and on the embodiment of ethical values (like transparency) in science policy.

[Zollman \(2007, 2010\)](#) looks at how network structure in a bandit model influences whether communities settle on correct or incorrect consensus. His surprising result is that across many variations of the model, groups perform better the less they communicate. Subsequently it has become clear that a host of models, with different varieties of problems, structures, and assumptions, yield similar results ([Fang, Lee, and Schilling 2010](#); [Frey and Šešelja 2018, 2020](#); [Grim et al. 2013](#); [Lazer and Friedman 2007](#); [March 1991](#); [Rosenstock, Bruner, and O'Connor 2017](#)).

These results go against the intuitive notion, underlying the open science movement, that more communication is generally better for science. But although initially surprising, once understood the result is intuitive. As noted, mechanisms that promote a period of diverse practice—agents testing multiple theories of some length of time—improve community output in these models. When agents communicate too much, they tend to settle on consensus too quickly, without a sufficient period of exploration. With respect to the open science movement, there is thus a risk in promoting sharing—that it might flatten the diversity of beliefs within a community and thus reduce diversity of practice.

To test this risk more explicitly, [Torsell and Korf \(2024\)](#) develop a different sort of network model. Agents explore an NK-landscape with “peaks” of different magnitudes corresponding to levels of epistemic success. Agents communicate with network neighbors about their current success levels and tend to follow neighbors who are more successful than themselves while also exploring locally. Torsell and Korf find that communities who share information more frequently can lead the entire group to herd onto local optima but miss better solutions because they do not explore enough. Thus, their results, again, suggest that too much sharing in science may inhibit progress by reducing diversity of practice.

It should be noted that there may be effective ways to both embody values related to openness and sharing in scientific communities and simultaneously protect diversity of practice. [Wu and O'Connor \(2023\)](#) point to areas of science that purposefully isolate research teams for periods of time to improve outcomes. In this way open sharing might be preserved as a general value in science, while in specific instances it could be limited to improve diversity of practice.

28.3.4 Public communication

The values in science literature has argued that scientists have ethical and epistemic responsibilities to publicly communicate their findings. Good communication by scientists may improve science literacy, reduce scientific misinformation, and counteract distrust in science ([Desmond 2023](#); [Gerken 2018, 2022](#); [Intemann 2023, 2024](#); [John 2018](#); [Kitcher 2011](#)).

There are a few models relevant to this topic. The first center around industry attempts to control scientific narratives. Both [Weatherall, O'Connor, and Bruner \(2020\)](#) and [Lewandowsky et al. \(2019\)](#) consider industrial agents who select what to share from real, independently produced data and, in doing so, manage to mislead the public. Neither of these models involves fraud or lying, but rather curation. Agents who wish to mislead curate which pieces of evidence the public sees and, in doing so, dramatically reshape public beliefs.

Another relevant model considers how this happens not with industry, but through channels of science journalism. [Mohseni, O'Connor, and Weatherall \(2022\)](#) consider a public receiving news items that are filtered through a journalistic process. Two communication strategies they consider are hyperbole—where journalists take real events and exaggerate them to sound more extreme or novel—and extremity—where journalists communicate accurately about events but select only extreme ones to share. They find that extremity—a form of curation—can be just as misleading as hyperbolizing, even though it involves sharing (just some of the) facts accurately.

There are a few take-aways for the values in science literature. First, ethical standards focusing primarily on truthfulness and accuracy of communication can be insufficient. Which facts are shared are often influenced by our values. This value-laden curation of accurate facts can, and very often does, mislead. Thus, those worried about public belief should focus, at least to some degree, on how scientific sharing is curated rather than primarily on reducing falsehood.

Second, as noted, there has been a focus on individual responsibilities of scientists to share their work in useful ways. This responsibility is undeniable. But attention to the emergent effects of industry strategy in science communication points to places where even the best individual communication will be insufficient. These strategies are often subtle, successful, well-funded, and carried out by professionals. Thus, it may be impractical for individual scientists to counter such efforts. A bigger picture should also address wider media and government policies that ensure successful public communication of science.

The main take-away for the values in science literature here is the suggestion that attention be given to a wider set of emergent social phenomena in thinking about science communication. It is not that these phenomena have been entirely ignored, but under-discussed. As a result, more focus is needed on the responsibilities of policies makers, rather than individual scientists, to protect public belief.

28.3.5 Out-group mistrust

Two recent models present surprising results about how individual prejudices can influence outcomes in epistemic groups. Both show that in some cases ignoring or devaluing data shared by outsiders can improve outcomes for some, or all, community members. These results are deeply connected to

claims from standpoint epistemology and epistemic injustice—two areas of thought with longstanding connections to values in science.

Wu (2023) starts with two communities in a network facing a two-armed bandit problem. She assumes that one of these—the dominant community—can entirely ignore or else partially devalue the data gathered by the other, marginalized, community.³ Across many versions of this model Wu finds that the marginalized community tends to arrive at accurate beliefs more often and also, typically, learns faster. In fact, they typically do better even than an entirely networked community. When a dominant group ignores data from a marginalized group in this model, they tend to learn more slowly and explore for a longer period of time. The marginalized group then learns more quickly but at the same time benefits from exploration happening in the dominant group. In other words, they are able to exploit the best current evidence because they do not ignore data, without risking a premature, incorrect consensus. Wu (2023) takes this model to provide a causal pathway for what is sometimes called the “inversion thesis” in standpoint epistemology—that those who are disempowered or socially marginalized are sometimes, paradoxically, in a better position to learn about certain aspects of the world (Wylie 2003).

Fazelpour and Steel (2022) consider a similar model, but where two groups both devalue data coming from their out-group. In an, again, somewhat surprising finding, they show that moderate out-group mistrust allows the whole group to perform better. This is, again, because mistrust slows learning and promotes exploration within the entire community. They also explore versions of the model where agents tend to conform with neighbors—a tendency that generally harms learners in these models (Mohseni and Williams 2021; Weatherall and O’Connor 2021)—with similar findings.

None of these authors advocate that scientists should cultivate mistrust toward identity-based out-groups. Rather both take their models as further support of arguments—from feminist philosophy of science and values in science—in favor of identity diversity in scientific communities (Longino 1990; Okruhlik 1994). Wu points out that including marginalized groups within science may mean including insights that benefit from historical exclusion. And Fazelpour and Steel argue this sort of diversity may benefit learning groups directly by decreasing herding onto sub-optimal theories. In this way, both of these models inform values in science.

28.3.6 Industrial withholding

Wu (2022) uses a version of the model just described to think about a quite different topic—the communist norm in science, which says that scientists should share their findings quickly, widely, and freely (Bright and Heesen 2023; Merton 1942). In particular, she is interested in the asymmetry between academic science, which adheres to the communist norm, and industrial science, which does not. Typically, industry science is proprietary, meaning that results are kept private and used for a profit rather than freely shared.

As Wu points out, though, this means that industry and academic science stand in a relationship much like the marginalized and dominant communities in her previous model. One group receives data from both communities, and the other can only learn from one. She develops a different version of the model where actors face an NK-landscape problem. The details are not important here, but the result supports her previous findings—if an industrial community withholds their data while simultaneously learning from an academic one, they learn faster and better. The academics, on the other hand, spend more time exploring possibly sub-optimal theories.

This points to a deep unfairness in the structure of research communities. Academic science is often publicly funded, but industrial science can use academic insights in developing new products and findings, which they do not need to then share. They, in turn, profit off these products and findings (O’Connor 2023). This asymmetry has not been widely discussed because it is not

necessarily obvious a priori, but the model developed by [Wu \(2022\)](#) makes clear why it should be worrying.

These results have obvious implications for the acceptability of certain types of value influence on the structure of epistemic communities. Industry science is aimed at profit—a very different value from those driving academic progress. While profit motives are useful and important to discovery in many cases, there is, as noted, an unfairness here, especially to members of the public whose tax dollars fund academic science. Thus, it might be appropriate to regulate the structure of industry science to ensure that industry does not unfairly profit from public funding.

28.4 Conclusion

As we have seen, interactions in learning groups—even simpler groups than those constituted by real humans—are often complex. Emergent phenomena can arise that are not easy to predict a priori. One important conclusion is that in thinking about normative standards for scientific communities, especially those crafted to promote scientific success, it is not always easy to assess which policies and norms will be best.

Here we have pointed to a number of network modeling results with interesting implications for the values in science literature—a literature deeply interested in the improvement of scientific functioning. We do not think that these models themselves should be generally taken to show what will happen in real groups of learners under analogous conditions. Rather they show some of the possible outcomes—including ones that sometimes support and sometimes press against central arguments and conclusions from the literature. These sorts of possibilities should be taken, at minimum, to complicate some theoretical arguments about science policy.

Some of these models, though, are further supported by examples, cases, or else empirical results that help validate their findings. [Holman and Bruner \(2017\)](#) argue that industrial selection actually occurred during the testing of anti-arrhythmic heart drugs to unfortunate effect. The “Zollman effect”—where less connection can mean more success—has been detected in a human subject experiment ([Derex and Boyd 2016](#)). As predicted by [Wu \(2023\)](#) and [Fazelpour and Steel \(2022\)](#) there is some evidence that in groups with heterogeneous identities there is a reduction of herding, with positive results ([Bear and Woolley 2011](#); [Phillips and Apfelbaum 2012](#); [Sommers 2006](#)). In each of these cases, there is more reason to think that the sorts of emergent phenomena identified in the models might be occurring in the real world and to take them seriously in the values in science literature.

To summarize, given its topic matters, the values in science literature should generally be sensitive to the possibility of emergent, surprising social phenomena. Models are one, but not the only, way to explore where and when these phenomena occur.

Acknowledgments

Thanks to Kaetlin Taylor for conversations about the relevance of network models to the open science movement. Many thanks to Kevin Elliott and Ted Richards for their work editing this volume.

Notes

- 1 See also Fernández Pinto and [Fernández Pinto \(2023\)](#) who confirm and extend this result.
- 2 Thanks to Kaetlin Taylor for suggesting the relevance of network models to the open science movement.
- 3 She directly relates this assumption to work on epistemic quieting ([Dotson 2011](#)) and testimonial injustice ([Fricker 2007](#)).

References

- Anderson, Elizabeth. 2020. "Feminist Epistemology and Philosophy of Science." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Spring 2020 edition. Stanford University.
- Bala, Venkatesh, and Sanjeev Goyal. 1998. "Learning from Neighbours." *The Review of Economic Studies* 65 (3): 595–621. <https://doi.org/10.1111/1467-937X.00059>.
- Bear, Julia B., and Anita Williams Woolley. 2011. "The Role of Gender in Team Collaboration and Performance." *Interdisciplinary Science Reviews* 36 (2): 146–53. <https://doi.org/10.1179/030801811X13013181961473>.
- Biddle, Justin B., and Anna Leuschner. 2015. "Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?" *European Journal for Philosophy of Science* 5: 261–78. <https://doi.org/10.1007/s13194-014-0101-x>.
- Biddle, Justin B., Ian James Kidd, and Anna Leuschner. 2017. "Epistemic Corruption and Manufactured Doubt: The Case of Climate Science." *Public Affairs Quarterly* 31 (3): 165–87. <https://doi.org/10.2307/44732791>.
- Bright, Liam Kofi, and Remco Heesen. 2023. "To Be Scientific Is to Be Communist." *Social Epistemology* 37 (3): 249–58. <https://doi.org/10.1080/02691728.2022.2156308>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Coates, Matthew. 2024. "Does It Harm Science to Suppress Dissenting Evidence?" *Philosophy of Science* 92 (1): 82–99. <https://doi.org/10.1017/psa.2024.21>.
- Cook, John. 2017. "Response by Cook to 'Beyond Counting Climate Consensus'." *Environmental Communication* 11 (6): 733–5. <https://doi.org/10.1080/17524032.2017.1377095>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2014. "Who's Afraid of Dissent? Addressing Concerns about Undermining Scientific Consensus in Public Policy Developments." *Perspectives on Science* 22 (4): 593–615. https://doi.org/10.1162/POSC_a_00151.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2018. *The Fight against Doubt: How to Bridge the Gap between Scientists and the Public*. Oxford University Press.
- Dere, Maxime, and Robert Boyd. 2016. "Partial Connectivity Increases Cultural Accumulation within Groups." *Proceedings of the National Academy of Sciences* 113 (11): 2982–7. <https://doi.org/10.1073/pnas.1518798113>.
- Desmond, Hugh. 2023. "The Ethics of Expert Communication." *Bioethics* 38 (1): 33–43. <https://doi.org/10.1111/bioe.13249>.
- Dotson, Kristie. 2011. "Tracking Epistemic Violence, Tracking Practices of Silencing." *Hypatia* 26 (2): 236–57. <https://doi.org/10.1111/j.1527-2001.2011.01177.x>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2018. "From Tapestry to Loom: Broadening the Perspective on Values in Science." *Philosophy, Theory, and Practice in Biology* 10 (8): 1–8. <http://dx.doi.org/10.3998/ptpbio.16039257.0010.008>.
- Douglas, Heather. 2021. *The Rightful Place of Science: Science, Values, and Democracy: The 2016 Descartes Lectures*. Edited by Ted Richards. Consortium for Science, Policy & Outcomes.
- Douglas, Heather, Kevin C. Elliott, Andrew Maynard, Paul Thompson, and Kyle Whyte. 2014. "Guidance on Funding from Industry." *SRPoiSE.org*. <http://srpoise.org/wp-content/uploads/2014/06/Guidance-on-Funding-from-Industry-Final.pdf>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Fang, Christina, Jeho Lee, and Melissa A. Schilling. 2010. "Balancing Exploration and Exploitation through Structural Design: The Isolation of Subgroups and Organizational Learning." *Organization Science* 21 (3): 625–42. <https://doi.org/10.1287/orsc.1090.0468>.
- Fazelpour, Sina, and Daniel Steel. 2022. "Diversity, Trust, and Conformity: A Simulation Study." *Philosophy of Science* 89 (2): 209–31. <https://doi.org/10.1017/psa.2021.25>.
- Fernández Pinto, Manuela, and Daniel Fernández Pinto. 2023. "Epistemic Diversity and Industrial Selection Bias." *Synthese* 201 (5): 182. <https://doi.org/10.1007/s11229-023-04158-7>.
- Frey, Daniel, and Dunja Šešelja. 2018. "What Is the Epistemic Function of Highly Idealized Agent-Based Models of Scientific Inquiry?" *Philosophy of the Social Sciences* 48 (4): 407–33. <https://doi.org/10.1177/0048393118767085>.
- Frey, Daniel, and Dunja Šešelja. 2020. "Robustness and Idealizations in Agent-Based Models of Scientific Interaction." *The British Journal for the Philosophy of Science* 71 (4): 1411–37. <https://doi.org/10.1093/bjps/axy039>.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

- Gerken, Mikkel. 2018. "Expert Trespassing Testimony and the Ethics of Science Communication." *Journal for General Philosophy of Science* 49: 299–318. <https://doi.org/10.1007/s10838-018-9416-1>.
- Gerken, Mikkel. 2022. *Scientific Testimony: Its Roles in Science and Society*. Oxford University Press.
- Grim, Patrick, Daniel J. Singer, Steven Fisher, et al. 2013. "Scientific Networks on Data Landscapes: Question Difficulty, Epistemic Success, and Convergence." *Episteme* 10 (4): 441–64. <https://doi.org/10.1017/epi.2013.36>.
- Holman, Bennett, and Justin P. Bruner. 2015. "The Problem of Intransigently Biased Agents." *Philosophy of Science* 82 (5): 956–68. <https://doi.org/10.1086/683344>.
- Holman, Bennett, and Justin P. Bruner. 2017. "Experimentation by Industrial Selection." *Philosophy of Science* 84 (5): 1008–19. <https://doi.org/10.1086/694037>.
- Intemann, Kristen. 2023. "Science Communication and Public Trust in Science." *Interdisciplinary Science Reviews* 48 (2): 350–65. <https://doi.org/10.1080/03080188.2022.2152244>.
- Intemann, Kristen. 2024. "Value Transparency and Promoting Warranted Trust in Science Communication." *Synthese* 203 (2): 42. <https://doi.org/10.1007/s11229-023-04471-1>.
- Intemann, Kristen, and Inmaculada de Melo-Martin. 2014. "Are There Limits to Scientists' Obligations to Seek and Engage Dissenters?" *Synthese* 191: 2751–65. <https://doi.org/10.1007/s11229-014-0414-5>.
- John, Stephen. 2018. "Epistemic Trust and the Ethics of Science Communication: Against Transparency, Openness, Sincerity and Honesty." *Social Epistemology* 32 (2): 75–87. <https://doi.org/10.1080/02691728.2017.1410864>.
- Kitcher, Philip. 2011. *Science in a Democratic Society*. Prometheus Books.
- Korf, Rebecca. 2025. "Values and Science Through a Social Lens." PhD diss., UC Irvine. ProQuest (Korf_uci_0030D_19537). <https://escholarship.org/uc/item/8s40w1q5>.
- Lazer, David, and Allan Friedman. 2007. "The Network Structure of Exploration and Exploitation." *Administrative Science Quarterly* 52 (4): 667–94. <https://doi.org/10.2189/asqu.52.4.667>.
- Leuschner, Anna. 2018. "Is It Appropriate to 'Target' Inappropriate Dissent? On the Normative Consequences of Climate Skepticism." *Synthese* 195: 1255–71. <https://doi.org/10.1007/s11229-016-1267-x>.
- Lewandowsky, Stephan, Toby D. Pilditch, Jens K. Madsen, Naomi Oreskes, and James S. Risbey. 2019. "Influence and Seepage: An Evidence-Resistant Minority Can Affect Public Opinion and Scientific Belief Formation." *Cognition* 188: 124–39. <https://doi.org/10.1016/j.cognition.2019.01.011>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71–87. <https://doi.org/10.1287/orsc.2.1.71>.
- Mayo-Wilson, Conor, Kevin J.S. Zollman, and David Danks. 2011. "The Independence Thesis: When Individual and Social Epistemology Diverge." *Philosophy of Science* 78 (4): 653–77. <https://doi.org/10.1086/661777>.
- Merton, Robert K. 1942. "A Note on Science and Democracy." *Journal of Legal and Political Sociology* 1 (1–2): 115–26.
- Mohseni, Aydin, Cailin O'Connor, and James Owen Weatherall. 2022. "The Best Paper You'll Read Today: Meida Biases and the Public Understanding." *Philosophical Topics* 50 (2): 127–53. <https://www.jstor.org/stable/48752503>.
- Mohseni, Aydin, and Cole Randall Williams. 2021. "Truth and Conformity on Networks." *Erkenntnis* 86: 1509–30. <https://doi.org/10.1007/s10670-019-00167-6>.
- Nash, Erin J. 2018. "In Defense of 'Targeting' Some Dissent about Science." *Perspectives on Science* 26 (3): 325–59. https://doi.org/10.1162/posc_a_00277.
- O'Connor, Cailin. 2023. "The Public Cost of Private Science." *Nautil.us*. <https://nautil.us/the-public-cost-of-private-science-312278/>.
- Okruhlik, Kathleen. 1994. "Gender and the Biological Sciences." *Canadian Journal of Philosophy* 24 (sup1): 21–42. <https://doi.org/10.1080/00455091.1994.10717393>.
- Oreskes, Naomi. 2017. "Response by Oreskes to 'Beyond Counting Climate Consensus'." *Environmental Communication* 11 (6): 731–2. <https://doi.org/10.1080/17524032.2017.1377094>.
- Paetkau, Tyler. 2024. "Inductive risk and Epistemically Detrimental Dissent in Policy-Relevant Science." *European Journal for Philosophy of Science* 14 (1). <https://doi.org/10.1007/s13194-023-00565-2>.
- Phillips, Katherine W., and Evan P. Apfelbaum. 2012. "Delusions of Homogeneity? Reinterpreting the Effects of Group Diversity." In *Looking Back, Moving Forward: A Review of Group and Team-Based Research*, edited by Margaret A. Neale and Elizabeth A. Mannix. Emerald Group Publishing Limited. [https://doi.org/10.1108/S1534-0856\(2012\)0000015011](https://doi.org/10.1108/S1534-0856(2012)0000015011).
- Rolin, Kristina. 2015. "Values in Science: The Case of Scientific Collaboration." *Philosophy of Science* 82 (2): 157–77. <https://doi.org/10.1086/680522>.

- Rosenstock, Sarita, Justin Bruner, and Cailin O'Connor. 2017. "In Epistemic Networks, Is Less Really More?" *Philosophy of Science* 84 (2): 234–52. <https://doi.org/10.1086/690717>.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgements." *Philosophy of Science* 20: 1–6. <https://doi.org/10.1086/287231>.
- Solomon, Miriam. 2001. *Social Empiricism*. MIT Press.
- Sommers, Samuel R. 2006. "On Racial Diversity and Group Decision Making: Identifying Multiple Effects of Racial Composition on Jury Deliberations." *Journal of Personality and Social Psychology* 90 (4): 597. <https://psycnet.apa.org/doi/10.1037/0022-3514.90.4.597>.
- Steel, Daniel. 2017. "Qualified Epistemic Priority: Comparing Two Approaches to Values in Science." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Torsell, Christian, and Rebecca Korf. 2024. "Network Study on Open Sharing." Unpublished.
- Weatherall, James Owen, and Cailin O'Connor. 2021. "Conformity in Scientific Networks." *Synthese* 198 (8): 7257–78. <https://doi.org/10.1007/s11229-019-02520-2>.
- Weatherall, James Owen, Cailin O'Connor, and Justin P. Bruner. 2020. "How to Beat Science and Influence People: Policymakers and Propaganda in Epistemic Networks." *The British Journal for the Philosophy of Science* 71 (4): 1157–86. <https://doi.org/10.1093/bjps/axy062>.
- Wu, Jingyi. 2022. "Withholding Knowledge." Unpublished. https://www.jingyiwu.org/uploads/1/3/8/0/138010336/wk_dec_11_2022.pdf
- Wu, Jingyi. 2023. "Epistemic Advantage on the Margin: A Network Standpoint Epistemology." *Philosophy and Phenomenological Research* 106 (3): 755–77. <https://doi.org/10.1111/phpr.12895>.
- Wu, Jingyi, and Cailin O'Connor. 2023. "How Should We Promote Transient Diversity in Science?" *Synthese* 201 (2): 37. <https://doi.org/10.1007/s11229-023-04037-1>.
- Wylie, Alison. 2003. "Why Standpoint Matters." In *Science and Other Cultures*, edited by Sandra Harding and Robert Figueroa. Routledge.
- Zollman, Kevin J.S. 2007. "The Communication Structure of Epistemic Communities." *Philosophy of Science* 74 (5): 574–87. <https://doi.org/10.1086/525605>.
- Zollman, Kevin J.S. 2009. "Optimal Publishing Strategies." *Episteme* 6 (2): 185–99. <https://doi.org/10.3366/E1742360009000063X>.
- Zollman, Kevin J.S. 2010. "The Epistemic Benefit of Transient Diversity." *Erkenntnis* 72 (1): 17–35. <https://doi.org/10.1007/s10670-009-9194-6>.

VALUING PARTNERSHIPS IN AGRICULTURAL GENOMICS RESEARCH

Using More Empirically Informed Philosophical Approaches to Advance the Science and Values Debates

Rachel A. Ankeny and Emily A. Buddle

29.1 Introduction

Although often bemoaned as a recent (and negative) development, academic and industrial research have long histories of interdependence (Parker et al. 2010; Swann 1988). Frameworks related to fostering scientific responsibility (Hackett 2016; Owen, Macnaghten, and Stilgoe 2012; von Schomberg 2013) can be viewed in part as responses to concerns about how scientists balance competing demands, including norms and drivers associated with partnering with industry. Copious scholarly literature on university-industry research partnerships¹ has documented problematic effects of industry involvement in academic science, particularly in the biomedical domain (Healy 2004; Sismondo 2009). Concerns range from how industry funding inappropriately biases research processes (Biscotti et al. 2009), including problem choice and priority setting, to suppression of results to suit commercial interests (Bravo, 2026; Elliott 2010; Fernández Pinto and Leuschner, 2026; Krinsky 2003; Sismondo 2008).

Many debates about these issues tend to view the values associated with industry (and other entities with economic motivations)² as in fundamental conflict with internal norms of academic science such as independence and freedom of inquiry, open flow of knowledge and findings, and the production of knowledge and associated benefits for the good of society. Values typically associated with publicly funded scientific research such as transparency, data sharing, and public benefit are viewed by some as in tension with values that underpin privately funded research, for instance, commercial benefits, shareholder obligations, and protection of data and intellectual property.

The reality is that university-industry research partnerships are ubiquitous in contemporary science (Ankrah and AL-Tabbaa 2015), particularly in Australian agricultural genomics research which is our focus. Drivers include rising costs and decreasing public funding, resulting in pressures on universities to collaborate with industry. There is increasing evidence in Australia and elsewhere that university-industry research partnerships help to drive the adoption of science and innovation in agriculture as they are argued to foster greater amounts of collaboration and pursuit of shared goals (Alston and Gray 2013; Paschen et al. 2021; Vasavi et al. 2025).³ There also are increasing pressures

on universities to fulfill broader social goals related to generation of useful knowledge and making documentable economic contributions to society, such as through entrepreneurship and the creation of academic “spin-offs” (Davey et al. 2024). University-industry research partnerships could be viewed as better at meeting these goals and providing considerable potential for enhancement of organizational capacity in contrast to more traditional research and development (R&D) processes.

We treat the question of what the potential benefits are for industry, academia, and other agricultural research-related parties in engaging in collaborative research as an open and empirical one that warrants more scholarly attention, particularly from philosophers of science. Many other types of scholars and policymakers recognize that industry and academic collaborations are critical for the types of innovation and knowledge exchange that are required to overcome the increasingly complex challenges that we face as a society (Abramo, D’Angelo, and Di Costa 2011; Cattaneo, Horta, and Meoli 2018). However, investigations have been limited perhaps due to facile assumptions about inherent conflicts particularly in scientific domains where big businesses are viewed as dominant such as in agricultural genomics (Ankeny, Bray, and McKinley 2018).

Hence in this chapter, we develop two lines of argument. First, we explore some features of university-industry research partnerships in a real-life example from agricultural sciences using a repertoires approach (Ankeny and Leonelli 2016) to problematize assumptions and claims regarding fundamental value conflicts associated with such partnerships. This exploration also shows how empirically rich philosophical approaches can be fruitful for advancing philosophical dialogues (Ankeny and Leonelli 2024a), and particularly about values in science, as they permit investigation of both the epistemic elements and non-epistemic factors associated with scientific practices without bracketing off the non-epistemic (Sample 2022). Such approaches allow us to engage more deeply with the range of normatively laden features of scientific practices in a specific context, which we contend can support further reflection on implications, for instance, for research training and governance, and science policy.

29.2 Background: Debates on science and values

Much existing scholarship begins by problematizing the supposed commonly held premise, drawing on long-standing analytic philosophy traditions, that science should aim to be free from all social, ethical, political, and other types of non-epistemic or contextual values as they bias or distort science (Douglas 2009; cf. Douglas 2014). More sophisticated explorations of these issues tend to focus on specific stages in science where values play significant (and potentially pernicious or inappropriate) roles, such as choices of research problems, methods, and aims, and how best to respond to evidence in terms of what inferences are supported by it, particularly for regulatory decision-making under uncertainty. There is a spectrum of which types of activities are included, depending on the commentator: For instance, Kevin Elliott (2017) uses a modestly expanded definition of scientific practices, including how to describe, frame, and communicate results. Nonetheless many such framings tend to exclude some activities as they are considered to be “outside of scientific research proper” (Douglas 2023, 252).

Our starting point differs from more traditional philosophical approaches to science and values: We explore a wide range of practices without predefining what counts as scientific research “proper.” We draw on approaches from philosophy of science in practice which emphasize that empirically informed philosophy can provide deeper access and insights into what occurs in science and serve as the basis for more robust philosophical analysis than can more traditional approaches (Ankeny and Leonelli 2024a, 2024b). These methods, including the analytical framework associated with the repertoires approach (Ankeny and Leonelli 2016), illustrate that it is impossible to strictly separate epistemic or cognitive factors from many non-epistemic or contextual ones (Longino 1996),

including those used to evaluate how research findings should shape decisions about applications or policy (sometimes called “practical” factors; see [Biddle \(2013\)](#)). Furthermore, it is important not to overidealize such practices as siloed from one another or immune to the influences of contextual factors but to excavate them in all of their complexity and messiness.

We have been informed by feminist philosophy of science scholarship, notably work by Helen [Longino \(1990, 2002\)](#), that explores the social factors at work in scientific communities which are associated with and structure the production of scientific knowledge. We widen our gaze to include the diverse factors associated with various communities or publics⁴ that contribute to the structures related to scientific knowledge production, given the deeply “hybrid” nature of much contemporary science, including the “technosciences” ([Sample 2022](#)) associated with contemporary agricultural sciences. Our approach is complementary to that defended by [Elliott \(2017\)](#): Given that value-free science is impossible (and not even an ideal toward which we should strive, as there can be legitimate goals that are value-laden), it is important to make values and their influences more transparent, and scrutinize them via robust engagement processes that include diverse interested parties.

Our approach takes account of additional criticisms raised about the typical framing of science in literature on values in science. For instance, Justin Biddle and Quill Kukla (2017) reject the “insistently discursive picture of epistemic activity” (217) present in the inductive risk literature, and its focus on scientific reasoning as seemingly performed by context-less individuals. Such a view overlooks a wide range of norms and understandings, particularly those that are collective or institutional, which are precisely the types of factors that we seek to include in our analysis. In applying the repertoires framework, we employ a more constructivist reading (informed in part by science and technology studies (STS) theories and approaches) of what counts as science as negotiated and achieved through the work of practitioners, including those formally considered scientists and others contributing to research such as farmers and breeders.

In agriculture, the concept of “values” is often tied to its economic meaning as a product-focused quantity related to market worth, a definition that carries over into the related sciences, for instance, in productive efficiency metrics to gauge likely research impacts. To give another example, moral philosophers’ arguments about giving non-human animals the same status or “value” as that of humans based on understandings of sentience ([Singer 1975](#)) are heavily contested among those in livestock production whose views in part include economic valuation ([Buddle, Bray, and Ankeny 2022](#)). In contrast, we use a common-sense, qualitative, and broad definition of “values,” namely anything that is viewed as worth pursuing or having by those who seek them, including both individual and more collective and institutionalized values, which is complementary to Elliott’s definition of a value broadly as “something that is desirable or worthy of pursuit” (2017, 11). We are well aware of the extensive debates over how we measure values, whether they are objective or subjective, if particular values are robust cross-culturally, and so on, including work in the philosophy of science that contends we should narrow our definition of values ([Biddle 2013](#)). For the purposes of this chapter, we focus not on conceptual clarifications but on illustrating how to make values and their influences (pernicious or benign) more visible and better understood in these contexts. We explore the practices not only of individual scientists but also of industry and institutional actors, including research development organizations, different levels of government, universities, and private industrial entities, via our experiences to date in a large cooperative research and training center focused on crop breeding.

29.3 Our setting: Research partnerships in the Australian landscape

Australia is an interesting case study for this exploration given the dominance of university-industry research partnerships ([Alston and Gray 2013](#)). Industry-funded research has increased significantly in recent years and is now the norm and expected of academics, particularly in agriculture where

industry-focused Research and Development Corporations (RDCs) contribute significant amounts of research funding ([Australian Government 2025](#)). These partnerships often bring with them expertise and technologies not available within university settings, while providing opportunities to make real-world contributions since research problems are co-constructed with specific benefits or impacts in view, such as in the Linkage Scheme of the Commonwealth Australian Research Council (ARC) ([Australian Government 2022](#)). In agricultural research, these partnerships often involve rural RDCs which collect levies from commodity producers and contributions from government departments and private industry, which in turn are matched with Australian government grant funding (up to 0.5% of the sector's gross domestic production) ([Australian Government 2025](#)). Contributions from RDCs include cash and in-kind such as materials, expertise, knowledge, staff time, infrastructure use, and/or laboratory access.

Other ARC funding schemes encourage academic researchers to collaborate with industry partners not only to leverage funds and conduct research aligned with industry priorities but to produce other benefits. The Industrial Transformation Training Centre (ITTC) scheme supports partnerships between university researchers and industry to provide bespoke training to higher-degree research students and postdoctoral researchers. These ITTCs offer PhD scholarships and postdoctoral positions typically aligned with an industry partner project, providing training opportunities in response to needs identified by industry and the researchers. We have participated in recent years in an ITTC, the Future Crops Training Centre (futurecropscentre.edu.au), which serves as the basis of our empirical work in this chapter.

29.4 Methods

Our empirical material is drawn from our observations and experiences in the Future Crops Training Centre, which focuses on the use of novel plant science techniques (particularly gene editing) to create new crop varieties that can thrive under changing environmental conditions. It is a partnership between two Australian Go8 universities, and over 20 research and industry partner organizations in Australia and overseas, together with 17 investigators, five postdocs, and 25 PhD researchers. It brings together diverse disciplines, including molecular biology, nanotechnology, crop physiology, plant science, microbiology, data science, science communication and rhetoric, social science, marketing, and philosophy working on theoretical, applied, and technological questions. The authors perform humanities and social science research and training in regulation, Responsible Research and Innovation (RRI), and science communication and public engagement for the Centre.

We use this ITTC as a rich example of a university-industry research partnership which provides a window into points at which values are in tension or shape research practices, rather than making any normative claims. Our data have been gathered from informal interactions during a range of types of administrative, governance, and strategic meetings, including in the process of applying for the ITTC; training sessions; center retreats; external workshops and conferences; PhD supervision activities; and discussions with our collaborators, including scientists, trainees, and industry partners.⁵ We draw on our own publications; academic and gray literature on genomic crop science; and background knowledge about regulation, the history of GM science and activism, and agricultural practices. As humanities and social science scholars who have performed significant amounts of public-private collaborative research,⁶ we are extremely well-positioned to share insights about how to study the framing and performance of industry-involved research, including against the backdrop of philosophical debates on science and values.

We utilize a repertoires framework to show that science is not just about theories and data but also about the habits, tools, values, and ways of working that scientists rely on every day to consider some of the key factors that we have observed, without developing an exhaustive analysis of the

scientific practices at work or making claims about whether the practices constitute a repertoire.⁷ The elements associated with a repertoire include conceptual and theoretical commitments, standards, infrastructures, communication strategies, social goals, funding, intellectual property, and institutional structures (Ankeny and Leonelli 2016). We have selected specific factors that we contend can ground a more refined view on university-industry research partnerships, as well as on the philosophical implications for science and values debates.

29.5 Factors associated with this university-industry partnership

R&D in agricultural genomics often involves longer-term investments, particularly in basic research particularly on plant traits hoped to be useful for future products but that are extremely distant from being market ready. The timeline for product development in GM research has been claimed by GM advocates to be extremely long, with a considerable amount of research required to develop and refine the underlying technologies (Agbio Investor 2022). Research in these highly technological fields requires the pooling of resources such as finances, equipment, and expertise, which allows cultivation of shared goals and even values over time. Furthermore, many successful university-industry partnerships are not one-offs but draw on long-standing collaborations and relationships (De Fuentes and Dutrénit 2012). In our ITTC, the university researchers and many of the industry partners collaborated on numerous research projects before the Centre commenced, grounded in previous professional relationships such as PhD supervision or working together in a laboratory. Grant writing took nearly two years, with close involvement of both the university researchers and industry partners, allowing careful articulation of goals, which in turn supported attention to shared values.

These details force us to rethink claims that industry partners will perturb or drive research priorities since they are solely or mainly focused on generating profit, hence making their values necessarily “pernicious” in terms of their effects on scientific practices. Some scholars have described how universities are reprioritizing revenue generation over the pursuit of knowledge as “academic capitalism” (Rhoades and Slaughter 2004). However, industry typically has its own pathways for pursuing highly applied research, either in house or by issuing competitive tenders with very clear guidelines, detailed expected outcomes, and rigid timelines; such projects are distinct from typical university-industry partnerships.

Furthermore, it is important to recognize that universities and scientists also invest in research, bringing their own expertise, time, and resources to the table when negotiating contributions for competitive grant applications. In the case of the ITTC scheme, the universities provided significant amounts of in-kind support, with cash granted by the ARC as a Commonwealth government entity. Hence resourcing is mutual, and no one party holds all of the power in relation to the financing and how it will be used, especially where multiple industry partners are involved as is required by the ITTC structures. This requirement helps to moderate the abilities of any one partner to drive the research toward a narrow focal area and fosters utilization of funding to benefit the goals of the broader center (e.g., to fund overarching social and ethical research components for which it is difficult to get individual industry partner support). The emphasis on training forces also consideration and development of arguments for how the capability building and research will benefit Australian society as a whole, both in the short- and long-run.

Industry partners typically fund and co-supervise one or more individual PhD projects (jointly with the university scientists), each of which does contain some commercial-in-confidence information and findings which are firewalled within project communications if required by IP agreements as determined before project commencement. However, these subprojects are part of the broader Centre research which can generate fundamental knowledge and novel technologies, and which in turn are shared through the usual academic channels such as publication and open data repositories. The

Centre's structure focused on training facilitates and requires interactions between the PhD students, as well as postdocs and industry and university researchers, which fosters communication and mutual learning and arguably reduces dominance of industry-specific values particularly where they are not aligned with broader Centre goals.

A closer analysis of this project's industry partners shows that a simple opposition between industry and public values cannot be assumed. Many of the partners are focused on developing certain types of crops as products, but some bring public-oriented values to the project, such as the RDCs which via their mandates must have their levy payers' benefits in view ([Australian Government 2025](#)). Such benefits are not merely economic but typically include considerations about the environment and rural communities, and what will benefit the sector as a whole. In this type of collaborative training center, industry partners need not all put in large amounts of cash. For example, advocacy organizations (e.g., Australian Seed Federation) provide broader contexts and perspectives, including exposure to research that is not university-driven. In general, Centre participants have shared motivations associated with participating for the "public good"⁸ or to be a "good citizen" in contributing R&D to their sector.

It is undeniable that there will be tensions between industry partners and university scientists (including investigators and trainees) in such partnerships that require careful navigation. However, we also have observed tensions in values among Centre researchers correlated with different methodological approaches (e.g., field versus laboratory research), diverse background disciplines (e.g., molecular biology versus agricultural science), those engaged with regulatory processes versus those working in more fundamental research (see [Ivanov, Buddle, and Ankeny 2025](#)), and those developing technologies versus working on specific crops, to highlight just a few examples. One worth unpacking in more detail illustrates different priorities among those trained and practicing as molecular biologists with passions for laboratory-based science, and those who consider themselves to be agricultural scientists (with backgrounds in fields such as crop science or plant genetics) who focus on research applications in the field or a particular industry context. There are clear differences in research priorities, and what is the best or most fruitful material to use, notably the default of using model plants such as *Arabidopsis* ([Provar, Alonso, and Assmann 2016](#)) as compared to work with important food crops but which have notoriously difficult genetics ([Anjanappa and Gruissem 2021](#)). Some critics in the broader scholarly literature have noted that reliance in much genomic-based plant research on model plants, together with performing experiments in highly controlled environments, is extremely problematic as they inadequately represent the full complexity of the biomes of interests, which in turn has created an innovation gap that has been very difficult to overcome ([Feuillet and Eversole 2023](#)). Although more holistic approaches are now being utilized, for instance, combining genetic and field technologies in ways that are more dynamic and responsive to more variable and changing environments (e.g., [Bowerman, Byrt, and Roy 2023](#)), disciplinary norms can be strong even in the most fruitful collaborative settings.

Broader and local institutional values also can lead to tensions or even conflicts in values. The Future Crops Training Centre is a collaboration between two Australian universities among the most prestigious and long-standing in the country, but which have very different values and norms that shape how research processes are valued and performed. The Australian National University (ANU) has particular strengths in fundamental biological research, whereas the University of Adelaide has extensive applied research programs especially in agricultural fields. These histories are associated with different internal metrics for evaluating success, for instance, related to scientific norms, including highly cited publications versus broader community engagement or even patents, as well as diverse attitudes toward valuation of resources (for a parallel example, see [Santos, Leach, and Ankeny 2025](#)). Although the structure of the Centre and the requirements of the underlying scheme

aim to reconcile such norms, there are inherent tensions and arguably conflicts in values that remain inescapable, and which we find ourselves making visible and navigating.

Positionality and experiences also contribute to value commitments: In agricultural genomics research, increasing production efficiency has been a traditional emphasis, particularly for food products aimed at lower-income countries. Recent technological developments notably via gene editing are often framed via this lens by both industry partners and some university researchers. However, this type of research tends to occur upstream and hence is extremely distant from its potential future applications, which then creates tensions with on-the-ground experiences of some PhDs who are from and have worked in the targeted low-income countries and tend to be motivated by concerns about social equity, food security, and the urgent need to combat climatic challenges at home. Another example can be found in the increasing emphasis in this field on environmental sustainability considerations, particularly in Australia which has been especially affected by the climate crisis: We have documented extensive use in the field of claims about sustainability with limited engagement around what sustainability actually means ([Wenzl, Buddle, and Ankeny 2025](#)), evidence of tensions in values associated with short- and long-term goals, as well as understandings of sustainability (e.g., economic productivity, environmental factors, and/or sociocultural issues). A final tension relates to the widespread neglect in crop science and related fields of fostering collaborations with Australian Indigenous communities but increasing recognition of the potential importance of native grains as genetic resources particularly given their adaptation to extreme climatic conditions.

These types of motivations, norms, and values not only reflect individual-level background, disciplinary, and career-stage differences, but broader geopolitical and socio-economic contexts that shape how individuals engage with science, and which are especially evident in agricultural genomics. Such internal diversity can be a strength for a research program such as ours, enabling richer dialogue and more reflexive research practices, including in relation to conflicts in values, but only if these tensions are explicitly acknowledged and carefully navigated. Managing these dynamics is essential to fostering inclusive research cultures and ensuring that collaborative research projects remain open to multiple ways of knowing and of valuing research and associated innovation processes.

This extended example underscores that not all values can be reconciled despite best intentions. The long history of debates over GM research, in part related to concerns about industry involvement, has led to strong opposition from some non-governmental organizations (NGOs) which tend to reject engagement with researchers not due to lack of interest but because they view their values as fundamentally in conflict. For example, Greenpeace has consistently opposed the development and use of genetically modified organisms in Australia (and elsewhere) and was associated with the destruction of a GM wheat field at a governmentally run experimental farm ([Nairn 2011](#)). Such actions have had long-standing impacts on scientists' views about whose voices and values should be included in research partnerships, due to their perceptions that value conflicts must be resolved in order to collaborate which is viewed as impossible in this case (for one potential approach to such conflicts, see [Ankeny 2025](#)). Involvement of relevant publics in discussions about agricultural genomics research continues to be a topic of active interest particularly in STS but warrants more attention by philosophers of science particularly those interested in policy (see [Ankeny 2016](#)).

29.6 Lessons learned for enriching the science and values debates

While there have been some explicit challenges in planning and conducting research within this Centre, the benefits have certainly outweighed the trials and tribulations. Not only has the blend of humanities and social sciences with biological and agricultural sciences enabled greater dialogue and consideration of phenomena beyond the lab, but the processes of navigating (but not necessarily resolving) tensions in values have been empowering for researchers, partners, and trainees.

One of the biggest challenges associated with recognizing values in real-time science is having occasions and space to stop and reflect. Unlike social science and humanities disciplines, reflexivity is not commonplace in scientific training or practice. The everyday pressures to produce fundamental knowledge and innovate often limit scientists' thinking to a particular experiment or lab; trainees typically land in preformed projects. Explicitly structuring projects to foster opportunities for researchers to identify and consider their blind spots or blinkers can in turn serve to generate more reflection about navigating contested values, whether it be between industry- and university-based researchers, different scientific disciplines, or those with diverse experiences and positionality.

A further finding derived from this university-industry partnership, and indeed collaborations across disciplines more generally, is that such experiences are not most usefully framed as creating spaces where pre-existing values need to be forced into alignment. Rather, these partnerships create spaces where activities of “valuing” happen, as seen in the constantly evolving, deeply situated, and often contested practices of deciding what matters, to whom, and why (on making the verb form “valuing” primary rather than the noun “value,” see [Kitcher \(2024, 232\)](#)). Focusing on valuing as an important activity, and indeed a scientific activity, shifts our attention away from static norms to prioritizing actions that arise through the negotiations and judgments that shape research pathways, priorities, and outcomes. This focus requires a more reflexive approach to partnership building, where tensions are not predefined (such as public versus private), nor assumed theoretically to be problematic. These activities undergird the generative part of the work of both building partnerships and doing research.

Recognizing university-industry partnerships as spaces in which valuing is enacted underscores the relational and context-dependent nature of research collaborations, that in turn require messy empirical investigations to adequately characterize and explore philosophically. As we have outlined in this chapter, the detailed contexts in which contemporary agricultural genomic science is practiced matter deeply because values are not only abstract ideals or primarily a theoretical matter but are experiential and can only be understood in the context of practices. To engage meaningfully in debates about values in science, we must attend to the full range of factors associated with scientific practices, including both epistemic and non-epistemic ones, and to the ways in which normative commitments shape and are shaped by the doing of science itself.

Acknowledgments

Both authors are affiliated with the ARC Training Centre for Future Crops Development (IC210100047) as part of Program 2: “Develop responsible research innovation practices” and receive research funding from it; Buddle receives partial salary support from the Centre. Ankeny is a member of the Gene Technology Ethics and Community Consultative Committee, an advisory committee associated with the Australian Commonwealth Gene Technology Regulator, and of the Social Science and Economics Advisory Group, Food Standards Australia New Zealand. Ankeny is a past member of the GM Crop Advisory Committee, Government of South Australia; Buddle is a current member. Ankeny acknowledges recent conversations with Hugh Desmond, Sabina Leonelli, and Helen Longino on science and values that contributed to this chapter, as well as hosting by the Department of Philosophy, Michigan State University in 2025.

Notes

- 1 We use the general terminology of “university-industry research partnerships” throughout this chapter while recognizing that various forms of partnerships may have different features; there is extensive literature on these types of partnerships of which we are aware but do not have space to explore in this chapter.

- 2 We define “industry” broadly to include any producers of goods or services that are not university-based, which includes not only private companies but also NGOs; Galleries, Libraries, Archives, and Museums (GLAM) sector organizations; and even governmental entities, in alignment with the Australian example analyzed in this chapter. It also is important to note that the boundaries between academia and industry have become more porous in recent years, especially in the medical and agricultural sciences, with the growth of research institutes and other entities that sit within universities but operate as businesses, as well as the cross-appointment of many academics to private companies, particularly in biotechnological fields.
- 3 There is an extensive literature within the philosophy of science and STS on collaboration which we recognize but do not have space to consider in detail (e.g., Andersen 2016; Genuth et al. 2000; Wagenknecht 2016); the metrics used by different disciplinary approaches to assessing the “success” of collaborations continue to warrant more philosophical analysis.
- 4 We use the term “publics” throughout this chapter in the sense in which it is used in STS to refer to the multiple and intersecting groups and communities to which people can belong, in contrast to any notion of a homogeneous general public or similar.
- 5 Ethics approval has been obtained from the University of Adelaide’s Human Research Ethics Committee (H-2023-142) and Australian National University’s Human Research Ethics Committee (H/2023/1122) to conduct empirical research on understandings of responsible research and innovation with Centre participants, including interviews and reflexive journals. However, no specific data collected under these approvals are used in this chapter.
- 6 Buddle’s PhD research was within an ARC Linkage Grant awarded to Ankeny and collaborators with industry partners (a major food retailer, boutique butcher, rural services organization, and state government R&D department) with matching contributions with public monies. Buddle is a farmer with work experience in agriculture policy at state and national peak body levels and has conducted contract research projects on livestock and crop production. Ankeny has a long-standing track record of collaborating with a range of industry partners on agriculture, food, and other topics, and has been a member of advisory committees focused on ethics and community involvement for the Australian Commonwealth Office of the Gene Technology Regulator (OGTR) for over 15 years.
- 7 We do suspect that contemporary genomic crop sciences may have a distinct repertoire, but the analysis is beyond the scope of this chapter.
- 8 The concept of the “public good” is philosophically and socially complex, and discussion of it goes well beyond the current chapter (e.g., see Roberson, Leach, and Raman 2021).

References

- Abramo, Giovanni, Ciriaco A. D’Angelo, and Flavia Di Costa. 2011. “University-Industry Research Collaboration: A Model to Assess University Capability.” *Higher Education* 62: 163–81. <https://doi.org/10.1007/s10734-010-9372-0>.
- Agbio Investor. 2022. “Time and Cost to Develop a New GM Trait: A Study on Behalf of Crop Life International.” Accessed December 23, 2025. <https://croplife.org/wp-content/uploads/2022/05/AgbioInvestor-Trait-RD-Branded-Report-Final-20220512.pdf>.
- Alston, Julian M., and Richard S. Gray. 2013. “Wheat Research Funding in Australia: The Rise of Public-Private-Producer Partnerships.” *EuroChoices* 12: 30–5. <https://doi.org/10.1111/1746-692X.12017>.
- Andersen, Hanne. 2016. “Collaboration, Interdisciplinarity, and the Epistemology of Contemporary Science.” *Studies in History and Philosophy of Science* 56: 1–10. <https://doi.org/10.1016/j.shpsa.2015.10.006>.
- Anjanappa, Ravi B., and Wilhelm Gruissem. 2021. “Current Progress and Challenges in Crop Genetic Transformation.” *Journal of Plant Physiology* 261: 153411. <https://doi.org/10.1016/j.jplph.2021.153411>.
- Anjanappa, Ravi B., and Wilhelm Gruissem. Forthcoming. “If Gene Editing Is the Solution, What Was the Question? Reframing Consideration of Gene-Editing Technologies in Agriculture.” In *The Social Epistemology of Engineered Agricultural Ecologies*, edited by Catherine Kendig and Paul B. Thompson. Springer.
- Ankeny, Rachel A. 2016. “Inviting Everyone to the Table: Strategies for More Effective and Legitimate Food Policy via Deliberative Approaches.” *Journal of Social Philosophy* 47: 10–24. <https://doi.org/10.1111/josp.12141>.
- Ankeny, Rachel A. 2025. “Reframing Gene Editing in Crops: Unpacking Potential Solutions by Reconsidering the Questions Asked.” In *The Social Epistemology of Engineered Agricultural Ecologies*, edited by Cathrine Kendig and Paul B. Thompson. Springer. https://doi.org/10.1007/978-3-032-04450-1_6.

- Ankeny, Rachel A., Heather J. Bray, and Kelly A. McKinley. 2018. "Collaborating with the Enemy? A View from Down Under on GM Research Partnerships." In *Ethical Tensions from New Technology: The Case of Agricultural Biotechnology*, edited by Harvey S. James Jr. CABI. <https://doi.org/10.1079/9781786394644.0053>.
- Ankeny, Rachel A., and Sabina Leonelli. 2016. "Repertoires: A Post-Kuhnian Perspective on Scientific Change and Collaborative Research." *Studies in History and Philosophy of Science* 60: 18–28. <https://doi.org/10.1016/j.shpsa.2016.08.003>.
- Ankeny, Rachel A., and Sabina Leonelli. 2024a. "Investigating Research Practices: How Qualitative Methods Enhance Philosophical Understandings of Science." *Qualitative Psychology* 11: 247–62. <https://doi.org/10.1037/qap0000289>.
- Ankeny, Rachel A., and Sabina Leonelli. 2024b. "Research Design for Philosophy of Science in Practice." In *Methods in the Philosophy of Science*, edited by Adrian Currie and Sophie Veigl. MIT Press.
- Ankrah, Samuel, and Omar AL-Tabbaa. 2015. "Universities-Industry Collaboration: A Systematic Review." *Scandinavian Journal of Management* 31: 387–408. <https://doi.org/10.1016/j.scaman.2015.02.003>.
- Australian Government. 2022. "Linkage Projects." Australian Research Council. Accessed May 27, 2025. <https://www.arc.gov.au/funding-research/funding-schemes/linkage-program/linkage-projects>.
- Australian Government. 2025. "Research and Development Corporations." Department of Agriculture, Fisheries and Forestry. Accessed June 4, 2025. https://www.agriculture.gov.au/agriculture-land/farm-food-drought/innovation/research_and_development_corporations_and_companies.
- Biddle, Justin. 2013. "State of the Field: Transient Underdetermination and Values in Science." *Studies in History and Philosophy of Science* 44: 124–33. <https://doi.org/10.1016/j.shpsa.2012.09.003>.
- Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190467715.003.0011>.
- Biscotti, Dina, Leland L. Glenna, William B. Lacy, and Rick Welsh. 2009. "The 'Independent' Investigator: How Academic Scientists Construct Their Professional Identity in University–Industry Agricultural Biotechnology Research Collaborations." *Economic Sociology of Work* 18: 261–85. [https://doi.org/10.1108/s0277-2833\(2009\)0000018013](https://doi.org/10.1108/s0277-2833(2009)0000018013).
- Bowerman, Andrew F., Caitlyn S. Byrt, Stuart J. Roy, et al. 2023. "Potential Abiotic Stress Targets for Modern Genetic Manipulation." *The Plant Cell* 35: 139–61. <https://doi.org/10.1093/plcell/koac327>.
- Bravo, Pedro. 2026. "Values and Industry-Funded Research." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Buddle, Emily A., Heather J. Bray, and Rachel A. Ankeny. 2022. "'Of Course We Care!': A Qualitative Exploration of Australian Livestock Producers' Understandings of Farm Animal Welfare Issues." *Journal of Rural Studies* 83: 50–9. <https://doi.org/10.1016/j.jrurstud.2021.02.024>.
- Cattaneo, Mattia, Hugo Horta, and Michele Meoli. 2018. "Dual Appointments and Research Collaborations Outside Academia: Evidence from the European Academic Population." *Studies in Higher Education* 44: 2066–80. <https://doi.org/10.1080/03075079.2018.1492534>.
- Davey, Todd, Sofia L. Martínez-Martínez, Rafael Ventura, and Victoria Galán-Muros. 2024. "The Creation of Academic Spin-Offs: University-Business Collaboration Matters." *The Journal of Technology Transfer*. <https://doi.org/10.1007/s10961-024-10153-y>.
- De Fuentes, Claudia, and Gabriella Dutrénit. 2012. "Best Channels of Academia-Industry Interaction for Long-Term Benefit." *Research Policy* 41: 1666–82. <https://doi.org/10.1016/j.respol.2012.03.026>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2014. "The Moral Terrain of Science." *Erkenntnis* 79: 961–79. <https://doi.org/10.1007/s10670-013-9538-0>.
- Douglas, Heather. 2023. "The Importance of Values for Science." *Interdisciplinary Science Reviews* 48: 251–63. <https://doi.org/10.1080/03080188.2023.2191559>.
- Elliott, Carl. 2010. *White Coat, Black Hat: Adventures on the Dark Side of Medicine*. Beacon.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Fernández Pinto, Manuela, and Anna Leuschner. 2026. "Epistemic Intimidation and Illegitimate Value-Influences in Science." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Feuillet, Catherine, and Kellye Eversole. 2023. "An Integrated, Systems-Wide Approach Is Needed for Public-Private Partnerships to Drive Genetic Innovation in Crops." *PLoS Biology* 21: e3002181. <https://doi.org/10.1371/journal.pbio.3002181>.

- Genuth, Joel, Ivan Chompalov, and Wesley Shrum. 2000. "How Experiments Begin: The Formation of Scientific Collaborations." *Minerva* 38: 311–48. <https://doi.org/10.1023/A:1026573717027>.
- Hackett, Edward J. 2016. "Essential Tensions: Identity, Control and Risk in Research." *Social Studies of Science* 35: 787–826. <https://doi.org/10.1177/0306312705056045>.
- Healy, David. 2004. *Let Them Eat Prozac: The Unhealthy Relationship Between the Pharmaceutical Industry and Depression*. New York University Press.
- Ivanov, Michail, Emily A. Buddle, and Rachel A. Ankeny. 2025. "Regulation as Key to Fulfilling the Promises of Agricultural Genomics: Going beyond Bottlenecks in Plant Gene Technology Development." *The Plant Journal* 122 (6): e70277. <https://doi.org/10.1111/tpj.70277>.
- Kitcher, Philip. 2024. "The Scientist, *Qua* Scientist, Is an Ethical Agent." *Filozofia* 79: 231–43. <https://doi.org/10.31577/filozofia.2024.79.3.1>.
- Krimsky, Sheldon. 2003. *Science in the Private Interest: Has the Lure of Profits Corrupted Biomedical Research?* Rowman & Littlefield.
- Longino, Helen. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.
- Longino, Helen. 1996. "Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy." In *Feminism, Science, and the Philosophy of Science*, edited by Lynn H. Nelson and James Nelson. Kluwer.
- Longino, Helen. 2002. *The Fate of Knowledge*. Princeton University Press.
- Nairn, Jessica. 2011. "Greenpeace Destroys GM Wheat." *ABC News*, July 14. <https://www.abc.net.au/news/2011-07-14/20110714-greenpeace-gm-protest/2794272>.
- Owen, Richard, Phil Macnaghten, and Jack Stilgoe. 2012. "Responsible Research and Innovation: From Science in Society to Science for Society, with Society." *Science and Public Policy* 39: 751–60. <https://doi.org/10.1093/scipol/scs093>.
- Parker, John N., Niki Vermeulen, and Bart Penders, eds. 2010. *Collaboration in the New Life Sciences*. Ashgate.
- Paschen, Jana-Axinja, Margaret Ayre, Barbara King, Nicole Reichelt, and Ruth Nettle. 2021. "Shaking It Up: The Realities of 'Doing' Co-Innovation in a Privatised Agricultural Advisory and Extension System." *Journal of Rural Studies* 87: 338–51. <https://doi.org/10.1016/j.jrurstud.2021.09.003>.
- Provart, Nicholas J., Jose Alonso, Sarah M. Assmann, et al. 2016. "50 Years of Arabidopsis Research: Highlights and Future Directions." *New Phytologist* 209: 921–44. <https://doi.org/10.1111/nph.13687>.
- Rhoades, Gary, and Sheila Slaughter. 2004. "Academic Capitalism in the New Economy: Challenges and Choices." *American Academic* 1: 37–59.
- Roberson, Tara, Joan Leach, and Sujatha Raman. 2021. "Talking about Public Good for the Second Quantum Revolution: Analysing Quantum Technology Narratives in the Context of National Strategies." *Quantum Science and Technology* 6: 025001. <https://doi.org/10.1088/2058-9565/abc5ab>.
- Sample, Matthew. 2022. "Science, Responsibility, and the Philosophical Imagination." *Synthese* 200: 79. <https://doi.org/10.1007/s11229-022-03612-2>.
- Santos, Dan, Joan Leach, and Rachel Ankeny. 2025. "Resources, Valuation, Trust: Sharing in Stem Cell Research." In *Science Communication and Trust*, edited by Antoinette Fage-Butler, Loni Ledderer, and Kristian H. Nielsen. Palgrave Macmillan Singapore. https://doi.org/10.1007/978-981-96-1289-5_17.
- Singer, Peter. 1975. *Animal Liberation: A New Ethics for Our Treatment of Animals*. Random House.
- Sismondo, Sergio. 2008. "Pharmaceutical Company Funding and Its Consequences: A Qualitative Systematic Review." *Contemporary Clinical Trials* 29: 109–13. <https://doi.org/10.1016/j.cct.2007.08.001>.
- Sismondo, Sergio. 2009. "Ghosts in the Machine: Publication Planning in the Medical Sciences." *Social Studies of Science* 39: 171–98. <https://doi.org/10.1177/0306312708101047>.
- Swann, John. 1988. *Academic Scientists and the Pharmaceutical Industry: Cooperative Research in Twentieth-Century America*. Johns Hopkins University Press.
- Vasavi, S., N. Anandaraja, P. P. Murugan, M. R. Latha, and R. Pangayar Selvi. 2025. "Challenges and Strategies of Resource Poor Farmers in Adoption of Innovative Farming Technologies: A Comprehensive Review." *Agricultural Systems* 27: 104355. <https://doi.org/10.1016/j.agry.2025.104355>.
- von Schomberg, Rene. 2013. "A Vision of Responsible Innovation." In *Responsible Innovation*, edited by Richard Owen, Maggy Heintz, and John Bessant. John Wiley.
- Wagenknecht, Susann. 2016. *A Social Epistemology of Research Groups*. Springer.
- Wenzl, Christina, Emily A. Buddle, and Rachel A. Ankeny. 2025. "Problematic Use of Sustainability Claims in Recent Scientific Literature on Crop Gene Technologies: Toward Improving Practices and Communication." *The Plant Journal* 122 (1): e70137. <https://www.doi.org/10.1111/tpj.70137>.

VALUES AND ASSESSMENT REPORTS ON CLIMATE CHANGE

Ahmad Elabbar

30.1 Introduction

How should we approach a “problem” like anthropogenic climate change? What exactly is the nature of its challenge, and what sort of responses might be appropriate? Despite what popular discussions might suggest, there are no easy answers to these questions. The causes of climate change go back centuries and implicate entire economic and political orders and ways of life. Its impacts, in turn, do not respect national borders, generational divides, or species boundaries, and will be experienced for centuries. And what it will take to mitigate or otherwise manage this unfolding crisis may require nothing short of changing the world.

Challenges on this scale have triggered innovations in the way in which scientific expertise is organized and delivered to policymakers. Once upon a time, scientific expertise entered public policy via intimate conversations among “eminent individuals” ([Oppenheimer et al. 2019](#), chap. 1). Picture a vizier leaning in and whispering in their sovereign’s ear a few words of wisdom, nudging them at the moment of action. Or think of a small council, convening on an ad hoc basis to address a well-defined question of public policy. Advisory mechanisms of this sort—small scale, responding to narrow requests for information, leaning heavily on individual savants—strain in the face of climate change. No individual or small group of individuals, no matter how eminent, can claim expertise on this challenge: The scope of interesting questions and relevant evidence is too vast, the range of disciplines is too numerous, the stakes are too high. Rather, circa 1970, the international community began instituting a web of “global environmental assessments” (GEAs) to comprehend a range of planetary crises and structure the flow of information between the scientific community and political institutions ([Jabbour and Flachsland 2017](#)). By now, “assessment” has become a technical term, picking out a novel site of knowledge production pioneered by GEAs ([Mach and Field 2017](#)). Their reports are social-epistemic projects, produced by massive groups of collaborating author teams, supported by an extended institutional bureaucracy. The most prominent of these GEAs is the Intergovernmental Panel on Climate Change (IPCC), established in 1988 and awarded the Nobel Peace Prize in 2007 for bringing the issue of climate change to public prominence.¹

The aim of this chapter is to survey the many ways in which social, ethical, and political values shape climate assessment reports, focusing on IPCC reports as paradigmatic examples. The IPCC has played a key role in contemporary debates over values in science, appealed to on all sides: By uncompromising defenders of the “value-free ideal” (e.g., [Betz 2013](#)) as well as their critics (e.g., [Steel](#)

2016), and many who hold middling positions in the debate (e.g., [Parker 2014](#)). Furthermore, beyond its relevance to the philosophical literature, understanding the value-ladenness of IPCC reports helps contextualize the broader discourse on climate change: Why the publication of an IPCC report, for example, is accompanied by both adulation and condemnation; why people disagree, even among those who fully accept the reality and severity of the climate crisis, on the merits of these reports.

The survey offered will be an opinionated one: In my view, the philosophical literature is currently limited by a near-singular focus on issues of reliability, uncertainty, and the social/political consequences of error in climate reports (see Elabbar forthcoming for a review). One would think, reading the extant literature, that the value-ladenness of assessment is a linear function of uncertainty: An “uncertainty monster” is purported to exist in climate science ([Curry and Webster 2011](#)), and the more we can “tame” that monster the less room there will be for values to impinge on climate reports. In the limit, if all the findings of climate reports could be made “error free,” the prospect of value-ladenness would disappear ([Betz 2017](#)). The chapter prompts philosophers to consider broader—and in many ways more entrenched—ways through which values shape climate reports, by drawing on the “IPCC studies” ([Hulme and De Pryck 2022](#)) literature across science and technology studies (STS), human geography, and political science, and suggesting ways their insights might be integrated into the philosophical literature on values in assessment.

To do so, the chapter introduces a “five features” framework of assessment, delineating a set of social-epistemic characteristics that distinguish assessments as novel sites of knowledge production, which are exemplified in the IPCC. These features are (1) the *institutionalization* of assessments; (2) their *political constitution*; (3) their *synthesis-orientation*; (4) their *curatorial nature*; and (5) their *agenda-setting power*. The value-ladenness of climate assessment reports cannot be analyzed adequately without recognizing these contextual features, as they constrain the agency of authors and the decisions that can be entertained in assessments, thus providing an institutional analysis of values in assessment that does not reduce to the psychological states or moral compasses of authoring scientists. The first two features will be discussed upfront, as they provide a “primer” on the IPCC and cut across all routes by which values shape assessments. The other three features will be threaded throughout the remaining sections, as they become relevant to more specific arguments.

30.2 A primer on IPCC assessments: Institutionalization and political constitution

Of the five features I will describe, perhaps the most conspicuous is the highly institutionalized nature of IPCC assessments—in at least two senses of “institution.”

First, the IPCC is an *organization*, with physical headquarters and venues ([Hartz and De Pryck 2022](#)), and an elaborate structure of subgroups and authorship hierarchies, reflected in the logos and official signatures at the outset of its reports that vouch for the organization’s “products.” This structure includes the Panel of UN Member States; the Bureau, which oversees the assessments; three Working Groups (WGs),² each responsible for a subset of the IPCC’s assessment mandate, delivering separate as well as joint reports³; a Task Force (TF) on National Greenhouse Gas Inventories; the Secretariat, and four Technical Support Units (TSUs), one attached to each WG and the TF, and play a central role in keeping the assessments on track and facilitating the work of authors ([Leclerc 2022](#); [Skodvin 2022](#)). Second, IPCC assessments are institutionalized in the sense of its epistemic practices being heavily *proceduralized* or *rule based*. While authors have considerable agency, they do not get to make whatever choices they deem appropriate. Rather, authorship in the IPCC is subject to a host of conventions, norms, and standards of best practice—including explicit procedures as well as “a list of unwritten rules” ([Yohe and Oppenheimer 2011](#), 633)—which serve to stabilize the organization’s outputs, imbuing its reports with an “institutional epistemology” ([Borie et al. 2021](#)) despite changes in its author teams.

Turning now to its *political constitution*, consider what gives IPCC reports their privileged standing in international climate politics. The IPCC is taken as the “epistemic authority” on climate change in the UN system not primarily on account of its standing within scientific communities, but rather because governments, among other political actors, play important roles in producing its reports. IPCC procedures carve out multiple points during its assessments for government participation, with the aim of securing shared ownership over its findings in both scientific and political fora (De Pryck 2022; Hughes 2022; Leclerc 2022). These moments occur at the outset of assessment, e.g., in the initial “scoping” phase, where the broad topics and outlines of reports are decided, and in the election of IPCC Bureau members⁴ and author nominations; as well as toward the end of assessment, in the reviewing of drafts and in the “acceptance,” “adoption,” and “approval” of different portions of the final reports—where “approval,” the most stringent category, denotes “detailed, line by line discussion and agreement” (IPCC 2013, 2) over the contents of the Summary for Policymakers (SPM), effectively giving governments veto power over this most influential portion of IPCC reports. As Skodvin (2022) notes, this intergovernmental design adopted for the IPCC at its founding in 1988 was “both novel and risky” and with “no precedent [...] in large-scale environmental assessments” (17). Explaining the rationale, or hope, that led its founders to trade in such control over authorship, leadership, and the contents of reports to governments, Howe (2014) puts it this way: “Once an inclusive IPCC released its conclusions [...] participating nations theoretically had little recourse to object about the basic facts when it came time to negotiate a framework convention on climate change” (159)—because these nations’ nominated scientists and delegates have signed off on the reports.

The reader may already have a sense of how these two features might have shaped IPCC assessments. To make their influence clearer, let us consider concrete, value-laden choices that arise in assessments and shape the contents of their reports, beginning narrowly with the problem of “inductive risk,” which has occupied the majority of the philosophical debate over values in the IPCC, and broadening from there.

30.3 The value-ladenness of climate assessment reports

30.3.1 Error and silence

The problem of inductive risk in assessment, and the ensuing debate over its implications for the IPCC, can be summarized as follows. Consider a group of authors confronting a complex body of evidence that bears on the truth of a hypothesis, *H*. The primary decision these authors face is whether to assess *H at all*. That is, whether the authors are willing to develop a “finding” that adjudicates *H*, despite the uncertainty in their judgments, or whether they deem it appropriate to suspend judgment and wait for more/better evidence. If they decide to suspend judgment, they are running “risks of silence”: The world will continue to change in the meantime, and withholding expert judgment for fear of error may undermine policymakers’ ability to respond adequately. On the other hand, if they decide to develop a finding, then they face a secondary decision regarding how to manage “risks of error”: If authors accept *H* and *H* turns out to be false (type I error), there are social and political risks being run, and the same is true if they reject *H* and *H* turns out to be true (type II error)—there are potential harms in basing climate policy on erroneous findings, either way.

Prima facie, it seems there is no *responsible* way of navigating these decisions without weighing the potential consequences of “silence” vs. “error” and of type I vs. type II errors—if authors choose to speak—and weighing these consequences appropriately requires making value judgments. Therefore, it seems to follow that values have essential roles to play in scientific assessments. At the end of a rather extensive debate over inductive risk in the IPCC (e.g., Betz 2013; John 2015; Steel 2016), the majority of philosophers have indeed drawn this conclusion and moved on to the more fine-grained

debate: Whether ethical and political considerations—of justice, autonomy, efficiency, etc.—favor managing inductive risk one way or another (Elabbar forthcoming).

It is important to note that, in practice, authors do not have full discretion over the management of inductive risk, even where they are aware and sensitive to the relevant trade-offs at stake—recall the “institutionalization” and “political constitution” of assessment. As such, the moral burden of managing inductive risk should not be understood as falling solely, and perhaps not even primarily, on authors’ shoulders. For example, the IPCC’s “Guidance Note” for lead authors on uncertainty directs them as follows: “Presentation of findings with ‘low’ and ‘very low’ confidence should be reserved for areas of major concern, and the reasons for their presentation should be carefully explained” (Mastrandrea et al. 2010). This deceptively simple sentence does two things: It favors “silence” as the default IPCC posture on highly uncertain matters but allows authors the discretion to develop “low” and “very low” confidence findings if the issue under assessment belongs to an area of “major concern.” A full article could be written on this fascinating passage. I refer to it in this discussion merely to emphasize that the management of inductive risk in the IPCC is an institutional matter, with both implicit and explicit conventions weighting the hands of authors. The political constitution of the IPCC also shapes authors’ decisions in the face of inductive risk. For example, in a penetrating study of uncertainty in the IPCC, Janzwood (2020) asks why “low certainty” findings “nearly disappear in executive summaries and SPMs” (1664), even though such findings—presumably judged as belonging to areas of “major concern”—are present in the underlying chapters. One interviewed author explains their exclusion from the SPM, with reference to the anticipated pushback from governments:

We only want the most robust messages in the SPM because we have to go through government approval—and if we have a whole bunch of low confidence statements in there, the governments can rightly ask, “if we have low confidence in this, why are you telling us about it? Tell us the things you know, not the things you don’t know.”

(Janzwood 2020, 1666–7)

The problem of inductive risk in the IPCC—and of managing value-laden decisions more broadly—should therefore be recast as an institutional challenge: Whether we agree or disagree with the IPCC’s posture on inductive risk, an adequate *explanation* of that posture should be given in terms of an institutional epistemology, focusing on the “epistemic capacities and dysfunctions of institutions” (Anderson 2025), rather than—or, at least, in addition to—the features of its individual, ever-changing authors.

30.3.2 Visual representation: Graphs, diagrams, and other “amplifiers of evidence”

IPCC visuals travel in ways their written findings do not, conveying information with an immediacy that is essential when communicating vast amounts of complex knowledge. IPCC authors and delegates know the power of imagery all too well, and some of the most serious controversies in the IPCC’s history have centered on visual representations. As a result, a huge amount of effort and scrutiny accompanies their production. Yet, while IPCC scholars from other fields have analyzed climate visuals and their controversies (Mahony and Hulme 2012; O’Reilly 2015; Walsh 2015), philosophers of science have rarely done so. The concept of “representational risk,” recently developed by Harvard and Winsberg (see Harvard and Winsberg 2026 in this *Handbook*), allows us to highlight one aspect of the value-ladenness of visual representation, but other philosophical ideas are needed to fully grasp the issue. Before going further, though, it might be helpful to encounter the moral dimensions of climate visualization at an intuitive level, through an example.

In 2019, the IPCC published its Special Report on the Ocean and Cryosphere in a Changing Climate (SROCC). In an underlying chapter, authors presented a striking figure showing the difference in global mean sea-level rise up to 2300 based on the near-term choice of emission pathways.⁶ The figure suggests that up to 5 meters of sea-level rise in 2300 might be avoided if policymakers were to take serious mitigation efforts in the short term. However, when deciding what information on future sea-level rise to include in the SPM (the only portion of the report likely to be read by policymakers), authors chose 2100 as the end-date, cutting the long-term projections. Presenting only the first portion of the figure up to 2100 risks limiting the horizon of policymaking to this date, and notably the stakes seem far less urgent on this horizon, as the projected sea-level rise for the two scenarios largely overlap until 2100, bifurcating significantly only beyond that date. Some government reviewers urged authors to redo the figure and include information beyond 2100, referring them back to their own underlying chapter and its longer time horizon.⁷ Ultimately, following further requests from governments during the live “approval” session of the SPM,⁸ the figure was updated to include the full projection up to 2300. This case further highlights the way the “political constitution” of assessment influences authorship.

An immediate refrain, when confronting this or similar cases where information is withheld, is to fall back on a presumed epistemic norm of full disclosure in scientific assessments. To insist that when reporting to policymakers, scientists should comprehensively disclose all policy-relevant information, that they should “not leave out relevant evidence” (Betz 2017, 101), and that this is what is troubling about the case—authors initially left out evidence when they ended the SPM figure at 2100. This response, however, is problematic in general, and particularly difficult to reconcile with the nature of visual (and other scientific) representations.

In general, as will be discussed further in a later section, a core feature of assessments is their “curatorial nature”—an epistemically successful assessment report is a curated report: Full disclosure of all relevant evidence on climate change would result in an unusable data dump. “Full disclose,” in short, is a false principle of advisory ethics and cannot be appealed to when criticizing the cases of withheld knowledge in assessment. If one objects to a specific finding being left out from an assessment report, or a salient portion of that report (say, the SPM), one must give particular reasons that justify its inclusion against the legitimate need to exclude vast swathes of information in assessment.

More specifically when it comes to visual representation, there are trade-offs between the amount of information packed into an image and the ease with which information relevant to a user can be discerned by that user (Bolinska 2016). This is something we are familiar with from overwhelming figures/charts that seek to represent every feature of the target system, at the cost of compromising epistemic access to any of its features. Again, if some disagree with SROCC’s authors and think they should have included more in their initial SPM figure (whether a longer-term trend, or additional scenario comparisons, rather just two extremes, etc.) that criticism cannot be made by appeal to a norm of full disclosure but must be sensitive to the inherently partial nature of representation. Using the language developed by Harvard and Winsberg—of “representational decisions” and their “adequacy-/inadequacy-for-purpose”—we might argue that the authors’ representational decision to exclude post-2100 sea-level rise rendered the figure “inadequate” for the purpose of reckoning with long-term impacts of present-day decisions and therefore involves taking significant “representational risks.” This, however, is no longer a purely epistemic objection: It appeals, implicitly, to views on climate ethics and intergenerational justice.

While Harvard and Winsberg’s work on representational decisions and risks helps reveal an aspect of the value-ladenness of visual representation, we need to go further to fully grasp the issue. Harvard and Winsberg’s notions capture trade-offs in the cognitive or epistemic aims of representation but do not speak to broader *affective* aims of visual communication. Visuals often aim at more than the mere conveyance of information: Images move us emotionally, capture and direct our imagination, and prompt moral reflection that is not easily afforded by other means of communication. For a salient example,

consider the prominent “climate generations” figure from the latest IPCC assessment (AR6).⁹ The figure—designed by communication experts in collaboration with authors—represents the prospects of a person born at different points in the twentieth and twenty-first centuries, from birth to old age, on different emission scenarios. The visual rhetoric of this image, more deliberately than the sea-level rise figure, draws the emotions of its audience and directly prompts moral reflection on the burdens borne by different generations.¹⁰ While philosophers have not engaged with these broader affective aims of visual communication, Paul Franco’s (2017, 2019) important effort to shift work on values-in-science in general away from its focus on cognitive content and attitudes toward the broader “perlocutionary effects” of speech acts—the full consequences of a communicative act on an audience—can be extended to imagery in science, as Franco himself suggests (Franco 2019, 1013, Footnote 5), and STS scholarship on the “visual rhetoric” in climate science is relevant here (Walsh 2015).

In summary, a broader philosophical engagement with climate visuals, attending to both cognitive and affective aims of visual communication, is still needed to grasp the ways in which values impinge on climate assessments, along the lines sketched in this section and beyond.

30.3.3 Conceptualization and the synthesis orientation of assessment

One of the enduring myths about assessments is that they produce no new knowledge or insights: That they are engaged merely in summarizing what is already known in the literature on climate change for the benefit of broader communities. There are many ways of seeing the inadequacy of this “summation” view of assessment (see Oppenheimer et al. 2019, chap. 6). One way is to recognize the extensive conceptual innovation that takes place in assessments: Some of the most influential concepts, schemas, and frameworks in climate research today first emerged, or were refined, in the drafting of IPCC reports. The IPCC’s hand in building the conceptual repertoire of climate science reveals another way in which values impinge on its assessments. Indeed, it is with respect to conceptualization that we get perhaps the most candid reflections on the role of values in the IPCC from its own authors. This is especially clear when authors develop “thick concepts”: Roughly, concepts that are neither purely descriptive nor purely evaluative, but which unite in their contents both elements of description and evaluation (see Alexandrova 2018).

To get an intuition for thick concepts, consider a news article that refers to a population of “vulnerable children” or “vulnerable young adults,” in some context. Perhaps a study has found that their number has increased as a result of economic policies. The idea behind thick concepts is that while empirical information guides the application of such concepts—facts about household income, medical record, educational attainment, etc., will be relevant to picking out “vulnerable children”—we need to make value judgments in addition: Concerning, say, the scope of harms that concern us, that we believe such children are unusually susceptible to, and the lives we fear they may lead as a result.

Turning to an actual example from the IPCC, consider the related concepts of “key [climate] vulnerabilities” and “key risks,” developed over multiple assessment cycles in WGII. In AR4, authors of the relevant chapter helpfully noted that their conceptualization of “key vulnerabilities” is value-laden in two ways. First, the criteria used to pick out “key vulnerabilities” include “normative elements” relating to “the uniqueness and importance of the threatened system, equity considerations regarding the distribution of impacts, the degree of risk aversion, and assumptions regarding the feasibility and effectiveness of potential adaptations” (IPCC 2007, 783–5). Second, even the choice of “which factual criteria to employ in assessing impacts has a normative component”—e.g., whether one takes the potential “persistence” of an impact to be relevant for conceptualizing “key vulnerabilities,” how one might operationalize “persistence,” and what weight to give it compared to other “factual criteria,” such as the “timing” of an impact, its “magnitude” or “scale,” etc.¹¹ Authors of the next iteration of this chapter in AR5 went further: The value judgments that are entangled in conceptualizing

key risks/vulnerabilities include the “importance of the vulnerable system,” including its “cultural importance” (IPCC 2014, 1052), and whether its persistence is “crucial for the survival of societies” (IPCC 2014, 1051), and whether it is possible to “replace such a system or compensate for potential and actual losses and damage” (IPCC 2014, 1051). As such, when introducing the final list of “key risks” drawn up in their chapter, authors emphasize that the list is “neither unique nor exhaustive: Other authors might express other preferences [...]” (IPCC 2014, 1068), depending on whether they endorse the relevant value judgments, among other sources of variability in expert judgment.

A reader may be feeling short-changed at this point. I began by stressing the inadequacy of the “summation view” of assessment by pointing to the novel concepts that emerge in assessment. I then suggested that, at least when such concepts are “thick,” values enter into assessment production via the route of conceptualization. But I have not explained why authors of assessment *must*, in some sense, create new concepts. Why can they not make do with the concepts already available in the literature—in line with the summation view—especially if creating novel concepts and conceptual frameworks might force assessors into the sort of difficult, value-laden decisions sketched in the example of key climate risks and vulnerabilities?

A full answer—not to the question of why scientists in general create concepts, but why scientists engaged in *assessments* do—awaits further research. But part of that answer, I suggest, will involve recognizing another feature of contemporary assessments: Their *synthesis-orientation*. Much of the underlying research on climate change will, appropriately, focus on advancing knowledge in a narrow area, and its conceptual repertoire will in turn reflect the needs of this narrower focus. Assessments seek the bigger picture: They exist not only to evaluate the merits of research efforts in subfields, but to combine insights from multiple areas of research into a synthesis with practical relevance for policymakers. To engage climate change *politically*, we typically require findings that are more informative with respect to practical action than those found in the underlying literature. We need to know, say, what the work of glaciologists on the fate of Antarctic glaciers implies for coastal planning, which requires combining their results with studies on urban development; regional projections of sea-level rise; the cost of adaptation measures in the exposed regions, how that cost compares to the cost of mitigation; etc. No doubt such synthesis efforts exist in the underlying literature. But assessments are unusually oriented toward producing findings that speak to the question: “how does it all fit together?”. Within the IPCC, this labor of synthesis takes two forms: What we might call *external synthesis* occurs when material from outside the IPCC is gathered and combined into basic findings in the underlying chapters. More distinctive of assessment is the work of *internal synthesis*: Where an assessment combines its *own* findings across chapters and WG reports. The higher up one moves in the IPCC’s synthesis pyramid, to ever-more exclusive summaries, the more intense this effort of internal synthesis, as evidenced by the extreme cross-referencing to chapters in the IPCC’s high-level findings reported in the SPMs. This synthesis orientation—eschewing first-order research¹² in favor of discerning what is already present, but implicit, in the underlying literature, producing novelty by connecting hitherto unconnected issues, and repackaging the insights of research into forms that might support practical action—requires a conceptual repertoire of its own. Hence the conceptual labor in assessment, and a route for values to impinge on assessment reports, where conceptualization is “thick” or, in other ways, involve contingent and morally consequential choices.

30.3.4 Attention cultivation: Epistemic curation and agenda-setting power

To close the chapter, I will sketch what I suspect is the most fundamental impingement of values on assessment, which also happens to be the least studied: The cultivation of attention.

When it comes to climate change, there is simply “too much to know”: Approximately 60,000 journal articles are published annually in climate science, to say nothing of the endless gray literature.

An assessment that sought to disclose everything implicit in this vast evidence base would deliver a useless data dump. To produce a report that is fit for *human* deliberation, under reasonable constraints of time and attention, the exclusion of vast amounts of relevant information is inescapable: Authors must decide what *not* to report. Furthermore, among what they do report, they must make finer judgments of priority: For example, which findings to elevate to the influential SPMs—the only pages likely to be read by policymakers. What I have elsewhere termed the “curatorial view of assessment” (Elabbar 2023) captures, among other things, the centrality of withholding knowledge in assessment: Of paring down vast epistemic worlds and discriminately selecting, at multiple stages, the issues deemed most urgent for an audience’s attention. This labor of selection and salience-making brings value judgments to the heart of assessment production in a more straightforward and entrenched manner than the other routes sketched in this chapter. That is, whatever one thinks of the significance of inductive risk, of representational choices, or of thick concepts, the basic, almost naïve, point—that assessment is selective, or, more broadly, curatorial in nature—clinches the case for the value-ladenness of assessment.

To appreciate the moral stakes of attention cultivation, via curatorial choice, it is important to highlight one more feature of contemporary GEAs: Their *agenda-setting power*.

The role of assessments in setting *political* agendas has been extensively studied. Indeed, much of the concern with the value-ladenness of IPCC reports arises from a keen appreciation of the structural power its reports exert on social discourse and policy. Less appreciated, but no less important, is the role of assessments in setting *research* agendas—hence I will focus on this second manifestation of agenda-setting power.

Scientific fields do not remain stable and unperturbed once “assessed,” but rather assessment reconfigures them, adjusting what is “considered interesting and worthwhile to research” in a particular field (Livingston and Rummukainen 2020, 14), “shaping agendas and funding, driving career paths, influencing publication strategies and citation patterns” (Hulme 2016, 223). While the evidence for the extent of this influence is still emerging, the studies to date are highly textured and compelling. For example, in a paper tellingly entitled, “Taking science by surprise,” Livingston and Rummukainen (2020) show how the IPCC’s decision to produce its now seminal “special report” on 1.5 degrees led to a reorganization of climate research efforts, prompting the broader climate science community to “engage with something that the majority previously saw as uninteresting and even unethical” (13). At the time of the IPCC’s decision to produce this report in 2016, only a handful of papers had been published on 1.5 degrees; yet, by the deadline for considering evidence for the IPCC’s special report, two years later a substantial evidence base of hundreds of papers on that temperature target had been produced at pace (Livingston and Rummukainen 2020, 13). A second example comes from Hughes and Paterson’s (2017) meticulous study of IPCC WGIII’s influence on climate mitigation research. They demonstrate that “key actors and institutions within this field are organizing their research, collaboration, and publication strategies around the assessment of knowledge in IPCC reports” (763), evidencing this claim by tracing the careers and institutional affiliations of an influential group of 44 highly connected authors, as well as through a broader institutional analysis of the field of climate mitigation research (749–62). Thus, they conclude:

It is hard to continue to think of the IPCC as an organization that is merely producing assessments of the state of current knowledge—as an organization separate from the fields of knowledge production that it is charged with assessing [...] the IPCC is defining both the terms of climate change mitigation knowledge production and global political action.

(Hughes and Paterson 2017, 763)

In summary, the curatorial nature of assessments, when coupled with their agenda-setting power, now in sharper view, creates a challenge which is at the very heart of the morality of assessment:

Which bits of the enormous challenge of climate change should we attend to at all? The curatorial choices made in influential assessments—regarding which evidence is to be assessed at all, which findings are worth communicating at all, and which findings are worth communicating *saliently*—are effectively interventions in an attention economy. They decisively shape the distribution of attention in both academic and political communities, thereby directing the actions of relevant agents and institutions toward particular aspects of climate change and away from others. The morality of this (necessary) exercise in attention cultivation is an open problem for the literature on values in assessment and remains largely unexplored.

30.4 Conclusion

This chapter surveyed the many ways in which values impinge on climate assessment reports. Through the balancing of the risks of error against silence, through choices of representation and conceptualization, and through the cultivation of attention, assessments may either protect or undermine the weighty interests at stake in climate action. In offering a broader range of arguments for the value-ladenness of assessment beyond those that draw on uncertainty and error, and in locating the site of managing values in assessment in institutions and their social-epistemic features rather than in the agency of individual scientists, this chapter hopes to encourage philosophers to ask new questions on the morality of assessment, and to engage the scholarship outside philosophy on the epistemic nature and moral stakes of climate assessment.

Notes

- 1 For a historical survey of its founding, see [Howe \(2014, 154–9\)](#).
- 2 On the “physical science basis” (WG1); “impacts, adaptation, and vulnerability” (WGII); and “mitigation of climate change” (WGIII).
- 3 Every five to seven years, the IPCC completes an “assessment cycle,” and the three WGs deliver individual reports as well as a joint “synthesis report.” There are also “special reports” produced in shorter timeframes addressing narrower topics (see [Livingston 2022](#)).
- 4 The “diplomat-scientists” who lead each assessment cycle ([De Pryck and Gaveau 2023, 76](#)).
- 5 Borrowing [Heekeren’s \(2021, 84\)](#) memorable phrase for science visuals.
- 6 [IPCC \(2019, fig. 4.2 on page 327\)](#), available through: https://www.ipcc.ch/site/assets/uploads/sites/3/2019/10/IPCC-SROCC-CH_4_2.jpg.
- 7 E.g., comment 1398 and 5456, available in “Government Review Comments” through <https://www.ipcc.ch/srocc/download/>.
- 8 See the minutes of the Earth Negotiations Bulletin on “Figure SPM.1” ([ENB 2019, 5](#)).
- 9 [IPCC \(2023, fig. SPM1\(c\)\)](#), available through: <https://www.ipcc.ch/report/ar6/syr/figures/figure-spm-1>.
- 10 As one of its designers reflects, “I have an emotional reaction when I see how my mother, myself, and my daughter are all living in very different worlds throughout our lifespans” ([Arlene Birt, quoted in Robaina 2023](#)).
- 11 Strictly speaking, it is only the first of these two considerations that renders a concept “thick”—the second, namely leeway in the choice of “factual criteria,” and the ethical significance of such leeway, renders it generically value-laden. I mention this second issue to briefly hint at another way in which even apparently technical, purely descriptive concepts can be value-laden ([Ohnesorge 2025](#)), and to highlight the degree of sophistication in these particular IPCC authors’ reflections on the role of values in science.
- 12 On the historical shift in assessments from producing original research toward synthesis of existing research, see [Oppenheimer et al. \(2019, 17, 195 and fig. 2.1\)](#).

References

- Alexandrova, Anna. 2018. “Can the Science of Well-Being Be Objective?” *The British Journal for the Philosophy of Science* 69 (2): 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Anderson, Elizabeth. 2025. “Local Knowledge in Institutional Epistemology.” *Australasian Philosophical Review* January: 1–23. <https://doi.org/10.1080/24740500.2024.2422551>.

- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3 (2): 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Betz, Gregor. 2017. "Why the Argument from Inductive Risk Doesn't Justify Incorporating Non-Epistemic Values in Scientific Reasoning." In *Current Controversies in Values and Science*, edited by Kevin C. Elliott and Daniel Steel. Routledge.
- Bolinska, Agnes. 2016. "Successful Visual Epistemic Representation." *Studies in History and Philosophy of Science Part A* 56: 153–60. <https://doi.org/10.1016/j.shpsa.2015.09.005>.
- Borie, Maud, Martin Mahony, Noam Obermeister, and Mike Hulme. 2021. "Knowing Like a Global Expert Organization: Comparative Insights from the IPCC and IPBES." *Global Environmental Change* 68: 102261. <https://doi.org/10.1016/j.gloenvcha.2021.102261>.
- Curry, Judith A., and Peter J. Webster. 2011. "Climate Science and the Uncertainty Monster." *Bulletin of the American Meteorological Society* 92 (12): 1667–82. <https://doi.org/10.1175/2011BAMS3139.1>.
- De Pryck, Kari. 2022. "Governmental Approval." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.025>.
- De Pryck, Kari, and Adèle Gaveau. 2023. "Scientists in Multilateral Diplomacy. The Case of the Members of the IPCC Bureau." *Political Anthropological Research on International Social Sciences (PARISS)* 4 (1): 65–105. <https://doi.org/10.1163/25903276-bja10040>.
- Elabbar, Ahmad. forthcoming. "Varying Evidential Standards as a Matter of Justice." *The British Journal for the Philosophy of Science*, August. <https://doi.org/10.1086/727429>.
- Elabbar, Ahmad. 2023. "The Curatorial View of Assessment and the Ethics of Scientific Advice: Beyond Decisional Autonomy towards Distributive Epistemic Justice." *IUHPST*. https://iuhpst.org/media/pdf/Elabbar_IUHPST_Prize_2023.pdf.
- ENB. 2019. "Summary of the 51st Session of the Intergovernmental Panel on Climate Change: 20–24 September 2019." *IISD* 12 (763): 1–21.
- Franco, Paul L. 2017. "Assertion, Nonepistemic Values, and Scientific Practice." *Philosophy of Science* 84 (1): 160–80. <https://doi.org/10.1086/688939>.
- Franco, Paul L. 2019. "Speech Act Theory and the Multiple Aims of Science." *Philosophy of Science* 86 (5): 1005–15. <https://doi.org/10.1086/705452>.
- Hartz, Friederike, and Kari De Pryck. 2022. "Venues." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.006>.
- Harvard, Stephanie, and Eric Winsberg. 2026. "Representational Risk Revisited." In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Heekeren, Simone. 2021. "Popular Science Images: Reflections on Visual Practices in Science Communication." *Design Issues* 37 (4): 72–85. https://doi.org/10.1162/desi_a_00659.
- Howe, Joshua P. 2014. *Behind the Curve: Science and the Politics of Global Warming*. University of Washington Press.
- Hughes, Hannah. 2022. "Governments." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.012>.
- Hughes, Hannah, and Matthew Paterson. 2017. "Narrowing the Climate Field: The Symbolic Power of Authors in the IPCC's Assessment of Mitigation." *Review of Policy Research* 34 (6): 744–66. <https://doi.org/10.1111/ropr.12255>.
- Hulme, Mike. 2016. "1.5°C and Climate Research after the Paris Agreement." *Nature Climate Change* 6 (3): 222–24. <https://doi.org/10.1038/nclimate2939>.
- Hulme, Mike, and Kari De Pryck. 2022. "Why the Need for This Book?" In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.002>.
- IPCC. 2007. Climate Change 2007: Impacts, Adaptation and Vulnerability. In *Contribution of WGII to the Fourth Assessment Report of the IPCC*. Cambridge University Press.
- IPCC. 2013. "Appendix A to the Principles Governing IPCC Work: Procedures for the Preparation, Review, Acceptance, Adoption, Approval and Publication of IPCC Reports." IPCC. <https://archive.ipcc.ch/pdf/ipcc-principles/ipcc-principles-appendix-a-final.pdf>.
- IPCC. 2014. Climate Change 2014: Impacts, Adaptation and Vulnerability. In *Contribution of WGII to the Fourth Assessment Report of the IPCC*. Cambridge University Press.
- IPCC. 2019. *The Ocean and Cryosphere in a Changing Climate: A Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.
- IPCC. 2023. *Climate Change 2023: Synthesis Report*. Intergovernmental Panel on Climate Change.

- Jabbour, Jason, and Christian Flachsland. 2017. "40 Years of Global Environmental Assessments: A Retrospective Analysis." *Environmental Science & Policy* 77: 193–202. <https://doi.org/10.1016/j.envsci.2017.05.001>.
- Janzwood, Scott. 2020. "Confident, Likely, or Both? The Implementation of the Uncertainty Language Framework in IPCC Special Reports." *Climatic Change* 162: 1655–75. <https://doi.org/10.1007/s10584-020-02746-x>.
- John, Stephen. 2015. "The Example of the IPCC Does Not Vindicate the Value Free Ideal: A Reply to Gregor Betz." *European Journal for Philosophy of Science* 5 (1): 1–13. <https://doi.org/10.1007/s13194-014-0095-4>.
- Leclerc, Olivier. 2022. "Procedures." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.005>.
- Livingston, Jasmine E. 2022. "Reports." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.007>.
- Livingston, Jasmine E., and Markku Rummukainen. 2020. "Taking Science by Surprise: The Knowledge Politics of the IPCC Special Report on 1.5 Degrees." *Environmental Science & Policy* 112: 10–6. <https://doi.org/10.1016/j.envsci.2020.05.020>.
- Mach, Katharine J., and Christopher B. Field. 2017. "Toward the Next Generation of Assessment." *Annual Review of Environment and Resources* 42 (1): 569–97. <https://doi.org/10.1146/annurev-environ-102016-061007>.
- Mahony, Martin, and Mike Hulme. 2012. "The Colour of Risk: An Exploration of the IPCC's 'Burning Embers' Diagram." *Spontaneous Generations: A Journal for the History and Philosophy of Science* 6 (1). <https://doi.org/10.4245/sponge.v6i1.16075>.
- Mastrandrea, Michael D., Chrisppher B. Field, Thomas F. Stocker, et al. 2010. "Guidance Note for Lead Authors of the IPPC Fifth Assessment Report on Consistent Treatment of Uncertainties." IPCC. <https://www.ipcc.ch/site/assets/uploads/2018/05/uncertainty-guidance-note.pdf>.
- O'Reilly, Jessica. 2015. "Glacial Dramas: Typos, Projections, and Peer Review in the Fourth Assessment of the Intergovernmental Panel on Climate Change." In *Climate Cultures*, edited by Jessica Barnes and Michael R. Dove. Yale University Press. <https://doi.org/10.12987/9780300213577-006>.
- Ohnesorge, Miguel. 2025. "We Should Not Align Quantitative Measures with Stakeholder Values." *Philosophy of Science*: 1–13. <https://doi.org/10.1017/psa.2025.10154>.
- Oppenheimer, Michael, Naomi Oreskes, Dale Jamieson, et al. 2019. *Discerning Experts: The Practices of Scientific Assessment for Environmental Policy*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226602158.001.0001>.
- Parker, Wendy. 2014. "Values and Uncertainties in Climate Prediction, Revisited." *Studies in History and Philosophy of Science Part A* 46: 24–30. <https://doi.org/10.1016/j.shpsa.2013.11.003>.
- Robaina, Eduardo. 2023. "We Spoke to the Person Who Developed the Climate Figure Everyone Is Talking About." *Climática*, April 12. <https://climatica.coop/climate-figure-ipcc-arlene-birt-ar6/>.
- Skodvin, Tora. 2022. "Origin and Design." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.004>.
- Steel, Daniel. 2016. "Climate Change and Second-Order Uncertainty: Defending a Generalized, Normative, and Structural Argument from Inductive Risk." *Perspectives on Science* 24 (6): 696–721. https://doi.org/10.1162/POSC_a_00229.
- Walsh, Lynda. 2015. "The Visual Rhetoric of Climate Change." *Wiley Interdisciplinary Reviews: Climate Change* 6 (4): 361–8. <https://doi.org/10.1002/wcc.342>.
- Yohe, Gary, and Michael Oppenheimer. 2011. "Evaluation, Characterization, and Communication of Uncertainty by the Intergovernmental Panel on Climate Change-an Introductory Essay." *Climatic Change* 108: 629. <https://doi.org/10.1007/s10584-011-0176-8>.

INDIGENOUS PEOPLES’ LEADERSHIP IN ENVIRONMENTAL SCIENCE ASSESSMENTS

Kyle Whyte and Pasang Yangjee Sherpa

31.1 Introduction

The Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) and the Intergovernmental Panel on Climate Change (IPCC) produce environmental science assessments that impact policymaking on environmental and economic issues that occur at the intersection of values and policy in science. For example, the IPCC reports shape public concern regarding values about risks of climate change. The concerns can shape major laws and policies in nations that affect governmental and philanthropic funding for renewable energy projects, hazard response infrastructure for extreme weather events, and the scientific and engineering research needed to advance such projects and infrastructure. As caretakers of millions of acres of terrestrial and aquatic ecosystems, many Indigenous Peoples throughout the world become concerned whenever a nation, or set of nations, decides to create an environmental science assessment on a topic related to biodiversity conservation, climate change mitigation, or food production. Indigenous Peoples have rights to exercise self-governance over conservation initiatives, food production and agriculture, and infrastructure related to energy and climate resilience. At the same time, they are also negatively impacted by the climate, energy, agricultural, and conservation projects of nation states, non-governmental organizations, and private industry, suffering excessive levels of pollution, violence and abuse, eviction from their territories, and unfair financial arrangements. The values that shape the scientific results of these reports can affect laws, policies, and programs that will have positive or negative impacts on Indigenous Peoples’ territories. Such laws, policies, and programs may not respect Indigenous Peoples’ self-governance rights to consent, self-determination, and decision-making in Indigenous territories.

We argue that Indigenous Peoples, whether as individual persons or as self-governing societies, ought to be equal leaders and participants in environmental science assessments sponsored by nations, both unilaterally and multilaterally. Additionally, we argue that Indigenous Peoples should have the freedom and support to organize and publish their own environmental science assessments using methods that arise from their communities’ questions, knowledge systems, and social situations. In this chapter, we will define some of the main reasons for concern about Indigenous leadership and participation in assessments and share some of the ways Indigenous Peoples have advocated to reform environmental science assessments and have endeavored to produce their own assessments.

31.2 Background

Indigenous Peoples are self-determining societies who have exercised self-governance and maintained territorial jurisdictions long before the establishment of nation states, including the members of the United Nations. There are nearly half a billion Indigenous Peoples in the world, with thousands of unique languages, cultures, histories, economies, and models of political organization. Groups who self-identify as Indigenous today are often those who—during the twentieth century—were not recognized or treated as holders of various human rights by nations, businesses, academic institutions, and non-profits. Indigenous Peoples suffered from rampant violations of their civil, political, economic, social, and cultural rights.

Ongoing rights violations have threatened Indigenous Peoples' responsibilities to protect the ecological conditions that support their communities' commercial and subsistence livelihoods, the sustenance of their cultures and societies, and their physical and mental health. For example, Native Hawaiians in Molokai, Hawaii, for generations, engaged in forestry, aquaculture, and well-organized subsistence hunting and fishing to maintain their economy. Yet, in the nineteenth and twentieth centuries, when the U.S. and state of Hawaii permitted American settlers to create ranches and resorts, these land-uses stripped the community of the ecological conditions for maintaining their sustenance, and the sacredness of places tied to their cultural integrity. Settler land-uses destroyed key biodiversity and blocked Indigenous sustenance practices, which have now set up the island for being vulnerable to climate change, especially drought, sea level rise, and coastal erosion. A group of Indigenous-led non-profit and community-based organizations have mounted a number of programs to defend their rights, fostering a movement around Aloha 'Āina, or "patriotic love for the land that feeds us" (Akutagawa 2019, 32). The organizations include, among others, 'Āina Momona, Sust'ainable Molokai, Ka Honu Momona, Ho'āhu Energy Cooperative Molokai, Hui o Kuapā, and Molokai Heritage Trust. They are leading to important conservation, livelihood sustenance, health, climate resilience, and energy outcomes. In 2025, the Molokai Heritage Trust, with support from the Global Land Restoration Fund, recently hosted the 'Āina Back Summit, an international conference which focused on "returning stewardship of 'āina to indigenous practitioners and communities across Hawai'i" and to "connect with initiatives and folks beyond our islands."¹

When we aggregate local actions like those of the community and community-based organizations in Molokai, the significance of Indigenous Peoples' conservation and climate mitigation activities are globally significant. Indigenous Peoples dwell on at least 28% of terrestrial areas, in about 87 countries (Garnett et al. 2018). Of lands in good ecological condition globally, 42% of them are found in Indigenous Peoples' territories (Forest Peoples Programme 2020). Overall, 36% of Indigenous territories overlap with key biodiversity areas (WWF et al. 2021). Indigenous territories are widely shown to be critical carbon sinks, including intact forests (Fa et al. 2020; Rights and Resources Initiative, Woodwell Climate Research Center, and Rainforest Foundation US 2021). A 2020 Report by the International Labour Organization found that "Globally, 18.2 per cent of indigenous peoples live below \$1.90 a day compared to 6.8 per cent of non-indigenous people" (ILO 2020).

Indigenous Peoples are rightly concerned, then, when nation states, individually or as groups, decide to launch environmental science assessments that will inform executives, legislators, and business and social leaders throughout the world about biodiversity, climate change, and food production. The mitigation of dangerous climate change effects and threats to biodiversity are major issues that many Indigenous Peoples tackle. As caretakers of ecosystems, Indigenous Peoples would not benefit from, and could be harmed by, any assessment that does not include Indigenous expertise and knowledge.

For example, the IPCC reports have offered analyses of models and interpretation of academic evidence, in various ways, that have led some people to support forest conservation programs, such as Reducing Emissions from Deforestation and Forest Degradation (REDD). Since IPCC's first featuring of such a program, REDD has, since then, had to be changed substantially. For REDD's approach

to forest conservation for climate mitigation threatened the forced displacement of some Indigenous Peoples, the destruction of Indigenous Peoples' customary occupations and livelihoods, and ignored the fact that Indigenous Peoples often were already managing forests successfully (Pathways Alliance for Change and Transformation (PACT) 2023). Had the IPCC included Indigenous expertise, knowledge, and voices, the character of its analyses and modeling, it is less likely that its audience in the policy world would have crafted forest conservation programs that threatened to harm some Indigenous People. We define Indigenous Knowledges as knowledges held by Indigenous Peoples based on their Indigenous worldviews, systems and values, and transmitted across generations. Indigenous Knowledges are peer-reviewed with regular assessment and evaluation (Whyte 2018).

Due to Indigenous Peoples' organizing around their concerns about the negative impacts of environmental science assessments on them, a number of such assessments have begun to develop methods for the fair and equal involvement of Indigenous Peoples and the inclusion of Indigenous Knowledge and issues (van Bavel, MacDonald, and Dorrough 2022). Consider two different assessments, IPBES and IPCC, which we will describe in what follows.

In 2012, IPBES was formed by 94 governments (and now includes 145 governments in 2025). IPBES' stated overall objective "is to strengthen the science-policy interface for the conservation and sustainable use of biodiversity, long-term human wellbeing and sustainable development," and would do so "by completing assessments on specific biodiversity themes" (e.g., pollination, invasive species) and methodological issues (e.g., values of nature and its benefits). In addition to its assessments, IPBES provides support on the development and use of policy-relevant tools and methodologies and builds "priority capacity, knowledge, and data needs" for IPBES "member States, experts, and stakeholders" (IPBES n.d.; UNESCO, IPBES, and UNU 2013). IPBES is currently beginning its second global assessment of biodiversity and ecosystem services.

The IPCC was created in 1988 by the United Nations Environment Program (UNEP) and the World Meteorological Organization (WMO) to assess the science related to climate change (IPCC n.d.). Their first Assessment Report (AR) was produced in 1990, to inform ongoing negotiations of the United Nations Framework Convention on Climate Change (UNFCCC). The assessment cycle begins with the election of the IPCC Bureau by the 195 Member Governments. The Bureau comprises 34 members who work to recruit authors and reviewers for the ARs. The three working groups (WGs) of the IPCC include the following: WG I The physical science basis; WG II Impacts, adaptation, and vulnerability; and WG III Mitigation of climate change. As of this writing, the IPCC seventh cycle is underway.

Science assessments, such as IPBES and IPCC, are institutions that are organized according to specific values that shape the character and credibility of their scientific outputs. IPCC and IPBES arose from the values embedded in major multilateral agreements, especially the Convention on Biodiversity, for IPBES, and the UNFCCC, for IPCC. The various values embedded in each assessment are clearly painted in the documents that govern the secretariat of each assessment: *The Principles Governing IPCC Work* and the *Functions, Operating Principles, and Institutional Arrangements of the IPBES*. The goals and text of these and other policies that shape IPBES and IPCC guide the purpose of the scientific assessments. The principles of the secretariat and the people who design the assessments use rules and procedures that govern the definitions of what counts as evidence, how scientific authors are selected, how ideas for assessments are vetted and approved, how drafts of chapters and final reports are determined, and the organization of events and conferences where scientists, diplomats, rightsholders, and stakeholders interact.

In each of these cases, values determine what methods, evidence, decision-making procedures, experts, and personnel are included or excluded. For anything or anyone that can be included, their inclusion may face greater or lesser barriers. For example, in IPBES and IPCC, nations are in the best position to nominate authors or people serving in other roles; at the same time, depending on the assessment, certain civil society organizations, and in various cases, Indigenous peoples, can make

nominations. In IPCC (and IPBES), diplomats, who represent the interests of their nations, have major say in the text of the reports. IPCC, for example, is understood by some ultimately publish results that privilege modeling methods from scientists from high-emitting countries and that favor allowing wealthy countries to keep higher energy consumption, including through using risky technologies through using risky technologies. Values about what types of expertise and evidence are relevant, how final approvals on scientific information are made, and which nations are positioned to have greater influence, are all ways in which values shape the policy-relevant scientific results of science assessments.

Indigenous scholars have long documented the way in which values in science shape assessments and the overall impacts that assessments can have on communities and national and global understanding of environmental issues. They have also focused on how values have excluded and shaped their involvement and the involvement of their knowledge. Andrea Tunks, for example, in 1997, writes about efforts in the late 1980s and 1990s of Māori leaders and knowledge holders questioning the values of the New Zealand Climate Change Programme, which produced scientific assessment reports and other documents during the 1990s. Those leaders and knowledge holders named values that excluded Māori knowledge, environmental ethics, and perspectives on climate change. They created a Māori Working Group in the Programme, which expressed values that recognized Māori knowledge and participation, and sought to elevate the climate issues faced uniquely by Māori communities (Tunks 1997). Tunks writes that “The purpose of the Māori Working Group was to ‘advise the Programme of Māori concerns and matters relevant to Māori and to ensure that the Programme is in accordance with its obligations under the Treaty of Waitangi’” (Tunks 1997, 87), which, critically, included Māori territorial self-governance. In stories analogous to that of the Māori Working Group, Indigenous Peoples across the world have worked to be involved in the beginning phases or just after the beginning phases of major science assessments, addressing values that excluded them and their knowledge from mattering. Indigenous Peoples are rarely included in the political actions or agreement proceedings that inaugurate the assessments, however.

Indigenous Peoples’ advocacy in science assessments was already shaped by many decades of Indigenous scholarship on science and values. For example, in the U.S., Indigenous scholars living at different times, such as Henry Roe Cloud (1884–1950), Ella Deloria (1889–1971), John Joseph Mathews (1894–1979), Vine Deloria Jr. (1933–2005), Cheryl Crazy Bull (living), and many others wrote and spoke about the way values affected research on Indigenous Peoples and participated in reforms to call out and change values that were negatively affecting Indigenous Peoples. They often called out the idea that dominant forms of science in the U.S. were covering up the influence of values in shaping scientific research and science education under the guise of “objectivity” or some notion of research absent of value influence. This was an observation that Indigenous Peoples readily could make, especially in cases where their own knowledge traditions are not ones that ever had tried to show that science was separate from society and societal values.

Scholars in fields that investigate the relationship of values and policy on science, including the philosophy of science and science and technologies studies, have tended to come much later to these conclusions, in the absence of having investigated more widely about the history of Indigenous Peoples and other groups who have been addressing values and science issues in dominant scientific institutions in the nineteenth and twentieth centuries. For example, in the philosophy of science, there have been active developments in the relationship between science and values (e.g., in the philosophy of science, see Brown 2020; Douglas 2009; Elliott 2017; Lacey 1999; Longino 1990). Yet the topic of scientific assessments has been a relative newcomer to this domain of the philosophy of science, and most writings are about IPCC (e.g., Elabbar 2023, 2025). Philosophers of science have tended to be more interested in critical analyses of how these assessments present value-laden assumptions about objectivity and the value-free ideal (Gundersen 2020; Jebeile 2022; John 2015, 2017; Petersen 2021)

or how the results of these assessments are strategically undermined by different types of skepticism (Biddle and Leuschner 2015).

More recently, some philosophers of science have begun to engage more with issues of scientific assessments and Indigenous Knowledge, especially emerging from the work of David Ludwig (Albuquerque et al. 2021; Ludwig 2016; Ludwig and El-Hani 2020; Ludwig et al. 2024; Ressorio et al. 2025), and with issues of equity and justice (Elabbar 2023, 2025) and Indigenous Peoples' knowledge (Massimi 2025). While some philosophers of science, such as Ludwig, have sought to learn from, reference, and build on the total array of scholars and intellectuals working on such topics, there is much opportunity for philosophers of science to really engage with the full extent of progress in understanding and intervening on issues of values and science in assessments. It is important that philosophers stand their definitions, concepts, and arguments on the shoulders of Indigenous and allied scholarship on Indigenous Knowledge, and the practical interventions that Indigenous Peoples have already experimented with. For example, in 2023 IPBES completed a report on its "Review of the inclusion of Indigenous and local knowledge in IPBES functions and deliverables," which offers clear and detailed findings about values, science, and policy.

The growth of literature on Indigenous Knowledge and science assessments could potentially dovetail with decades of scholarship and interventions by Indigenous Peoples in science assessment processes. So, the examples we describe here regarding IPBES and IPCC did not occur in a vacuum and are certainly not the result of articles and books published in the philosophy of science. They are building on longer and older dialogues on the exact issues that Indigenous Peoples have marshaled to ensure they have a meaningful place in science assessments. Consider, in more detail, a snapshot of Indigenous advocacy at the intersection of science and values in IPBES. We can start with the idea that "An IPBES assessment is a critical evaluation of the state of knowledge by independent experts interacting with government representatives and other stakeholders in a stepwise process to ensure legitimacy, relevance and credibility" (Baptiste 2020). At the very first plenary, in 2013, Joji Carino delivered the "Statement by Indigenous Peoples and Local Communities to the IPBES1 Plenary." The statement states how:

One of the biggest challenges relating to the loss of biodiversity and ecosystems today is the associated decline of indigenous, traditional and local knowledge: those uniquely adapted relationships and management systems which have sustained cultural landscapes and seascapes over millennia. The same drivers and underlying causes which result in biodiversity loss also underpin threats to diverse cultures—which are among humanity's richest and tested ways of caring for and sustaining the intrinsic values and ecosystem services from the natural world. IPBES needs indigenous and local knowledge and can support its dynamic revitalization and continuing relevance, more than ever today.

(Carino 2013)

The statement read by Carino (2013) goes on to describe a solution that will work to help ensure that positive outcomes will result from IPBES' assessments. The solution involves the greater participation of Indigenous Peoples:

IPBES needs to become a truly diverse knowledge platform, transcending sciences and embracing multiple expertise and knowledge holders. Such a platform cannot be operated as a top-down structure, driven by inflexible rules of procedure, but requires skillful and innovative partnerships with the existing knowledge networks from local-global scales, including those of Indigenous Peoples and local communities.

(Carino 2013)

The statement recommends interventions that will improve the IPBES assessments for the sake of Indigenous Peoples and their territories. They ensure that Indigenous Peoples will be respected as participants at different levels, that Indigenous persons will not be barred by discrimination against their credentials, that there will be funding to support their participation, that there should be safeguard policies protecting against the exploitation of Indigenous Knowledge, that mechanisms for inter-cultural collaboration will exist, and that there will be respect for rights, especially consent. The interventions include specific roles Indigenous expertise should play in key parts of the IPBES proceedings, and assurances that IPBES' results will benefit Indigenous Peoples' activities on the ground.

Consider, similarly, how Indigenous advocacy regarding values and science has worked for the IPCC. An IPCC assessment "identifies where there is agreement in the scientific community on topics related to climate change, and where further research is needed. The reports are drafted and reviewed in several stages, thus guaranteeing objectivity and transparency" ([Belgian Earth Observation 2021](#)). Again, IPCC reports inform ongoing negotiations of the United Nations Framework Convention on Climate Change (UNFCCC). Part of the efforts of Indigenous peoples to seek meaningful involvement of Indigenous Knowledge and issues in the IPCC has been through their advocacy in the UNFCCC.

As the first co-chair of the Local Communities and Indigenous Peoples' Platform (LCIPP) of the UNFCCC, Pasang Dolma Sherpa, executive director of the Center for Indigenous Peoples' Research and Development (CIPRED), stated that "We need the world to recognize the role of Indigenous People's knowledge, cultural values, and customary institutions in protecting the planet's ecosystems and climate" ([ThinkLandscape, n.d.](#)). Indigenous Elders who participated in the early climate change meetings advocated for Indigenous Peoples' rights, but they were hardly visible at that time, Pasang Dolma has noted. Reflecting on the changes in the UNFCCC, she described that "it took almost 20 years for the roles and contributions of Indigenous Peoples and local communities in climate change resilience to be recognized" ([Sherpa 2019](#)). She stated that

When the 2015 Paris agreement established the Local Communities and Indigenous Peoples Platform (LCIPP), for the first time in the UNFCCC's history since 1994, it gave Indigenous Peoples hope of having a platform to share their voices and be heard in climate change negotiations. At COP24 in Katowice, Poland in 2018, the UNFCCC then decided to have 14 members in the LCIPP, 7 appointed by Indigenous Peoples and 7 chosen by the national governments that are the parties to the UNFCCC. By giving equal representation, the decision at this COP legitimized the role of Indigenous Peoples and acknowledged that they should be respected.

(Goyal 2020)

Indigenous Peoples' experiences with IPBES and IPCC represent a broad array of Indigenous advocacy in environmental science assessments that affect values and policy in science. We argue that Indigenous Peoples, both as individuals and as self-determining societies, ought to be equal participants and leaders in environmental science assessments sponsored by nations, both unilaterally and multilaterally. Additionally, we argue that Indigenous Peoples should have the support to organize and publish their own environmental science assessments using methods that arise from their communities' questions, knowledge systems, and social situations.

31.3 Modes of exclusion of Indigenous Peoples from environmental science assessments

Earlier we defined Indigenous Knowledges as knowledges held by Indigenous Peoples based on their Indigenous worldviews, systems and values, and transmitted across generations. Indigenous Knowledges are peer-reviewed with regular assessment and evaluation. Over time, each Indigenous People

developed their own rigorous knowledge systems for guiding how their societies interacted with ecosystems in which they dwell. For example, some Indigenous Peoples of North America developed regenerative food systems through their management of bison. Bison grazing

improves the root structure of the grasses and soil health by, among other things, increasing the soil's retention of rainwater. Their shaggy coats distribute seeds across the landscape, and the wet spots where they wallow support birds and other species.

(Chesnais et al. 2024, 50)

Knowledge of how to care for bison and respect for bison is baked into family, social, cultural and ceremonial, and political life.

The knowledge systems of Indigenous Peoples are significant for assessing environmental problems pertaining to climate change, food production, and biodiversity conservation, and for making recommendations about how to solve them. Yet, when nations have sponsored environmental science assessments to better understand problems they are alarmed by, they have historically excluded Indigenous Peoples from participating in those assessments. Hence, for example, the first reports of a number of assessments, including the IPCC, had no Indigenous persons serving as authors, staff, or chairpersons. Moreover, Indigenous Knowledge systems were not referenced at all. Increased participation and support for Indigenous authors in the IPCC process is crucial (Carmona et al. 2023).

Indigenous Peoples' movements to participate in environmental science assessments have worked to end the assumption that Indigenous Peoples are *not* "knowers." They have shown how Indigenous rights, values, and knowledge can improve the quality and impact of scientific assessments, and what changes and transformations must be made within assessment processes. Second, they have begun the early stages of outlining what Indigenous-led, independent assessments would look like on topics such as biodiversity conservation, climate change mitigation, health equity, and food sovereignty.

Indigenous advocacy within multilateral scientific assessments is largely not understood outside of some of the assessment processes themselves. The history of Indigenous Peoples advocacy is also likely to be best understood as starting much earlier than the latter half of the twentieth century. Historically, many international agreements governed environmental-related issues and had something to do with knowledge of the environment thereof. Though, historically, these agreements seem little like IPBES, IPCC, and others today. In the Peace of Montreal, in 1701, a transatlantic agreement was made to regulate the fur trade. Diverse Indigenous Peoples of North America treated with New France. Many of the Indigenous delegates signed with symbols of the clans that they were from, representing their responsibilities to ecosystems in the areas they came from, as opposed to simply property or governance rights (Bohaker 2006).

In early to mid-twentieth century, conferences and assessments related to Indigenous Peoples began to emerge, such as the Inter-American Indian Conference, which was held in Pátzcuaro in 1940 and led to the creation of the Inter-American Indian Institute. The Institute, among other things, sought to conduct "Scientific investigations in regard to Indian problems" (United Nations (UN) 1982, 333). At the first conference, some Indigenous persons participated, but their own speeches were not allowed to be documented, even though those of non-Indigenous persons were (Kaltmeier 2016). The Meriam Report in 1928 was created by the U.S. to assess major issues affecting Indigenous Peoples, using a social science method. Henry Roe Cloud, an Indigenous person, served on the committee overseeing the report, and as an author on the report. He enacted his role strategically, serving to mitigate some of the racial bias, and ultimately ensured at least some of the results of the study would be relevant to the circumstances of Indigenous People in the U.S. sphere (Ramirez 2009).

While the aforementioned reports, again, are not precisely analogous to environmental science assessments today, it is nonetheless true that Indigenous advocacy today is shaped by the history of

Indigenous persons seeking to limit their exclusion from major agreements and assessments. Later in the twentieth century, Indigenous Peoples fought to be active in the emerging environmental science assessments. The minutes of the first [IPBES \(2013\)](#) plenary document show that

Ms. Cariño described the modes of participation for Indigenous Peoples in several different intergovernmental processes, including the Arctic Council and its Working Groups, where Indigenous Peoples sit as Permanent Observers alongside governments, and the Convention on Biological Diversity (CBD) and its Working Group on Article 8 (j) and related provisions, where Indigenous Peoples and governments participate in debates as equals. (14)

Since the early days of climate conferences, Indigenous leaders have advocated for greater presence in climate change negotiations ([Sherpa 2019](#)). Such advocacy can be found in International Indigenous environmental and climate change declarations that provided distinct understanding of Indigenous environmental justice ([McGregor 2020](#)). The *Kari-Oca Declaration*, for example, provided Indigenous Peoples' view on sustainable and humane life on Earth at the Earth Summit convened at Rio de Janeiro in 1992. However, Indigenous Peoples' contributions have not been recognized for the most part ([Ford et al. 2016](#)).

Indigenous Peoples' advocacy has historically faced barriers to their meaningful participation in the planning, decision-making, implementation, and reporting of these assessments ([Ford et al. 2016](#); [Ford, Vanderbilt, and Berrang-Ford 2012](#); [Smith and Sharp 2012](#); [Tengö et al. 2014, 2017](#); [van Bavel, MacDonald, and Dorrough 2022](#)). Indigenous scholars have questioned the ethnocentric and colonial assumptions made by scientists in assessments about the relationships among science, values, and policy, including assumptions that have objectified Indigenous Knowledge as merely raw data—as opposed to rigorous knowledge—and denied agency and expert credentialing to Indigenous Knowledge holders. Indigenous advocacy within multilateral science assessment processes has generated small wins, including increased references related to Indigenous Peoples and Indigenous Knowledge systems in the IPCC AR6 ([Carmona et al. 2023](#); [Sherpa 2025](#); [van Bavel, MacDonald, and Dorrough 2022](#)). However, most of the references are included in the WG II report on Adaptation, Vulnerability and Impact instead of the Physical Science basis that defines the “problem” of climate change for the assessment ([Carmona et al. 2023](#)).

Exclusion of Indigenous Peoples in global climate assessments is further explicated through a consideration of the unevenness of co-authoring relations ([Caretta and Morgan 2021](#)). The epistemic influence from a small number of co-authors co-writing regularly ([Corbera et al. 2016](#)) constitutes what we call the ecosystems of authorship—heavy overlaps in the authorship and journal publishers across scales resulting in the same ecosystem of individuals and teams, knowledge, and value being reproduced. These ecosystems of authorship produce Euro-Western science-based articles for select peer-reviewed journals based in Europe and North America, which gets fed into regional and global assessments. Such knowledge base restricts meaningful engagement with Indigenous Knowledge, especially when they do not appear in international peer-reviewed academic journals. When Indigenous Knowledge does appear in the literature, it has been shown that the engagement with Indigenous Knowledge and Indigenous Peoples are largely extractive in nature and significantly less are co-produced ([David-Chavez and Gavin 2018](#); [Klenk et al. 2017](#)).

31.4 Current interventions in major scientific assessments

IPBES and IPCC are not alone. There are many national and multilateral institutions that advance scientific assessments on environmental problems, globally, that are of concern to Indigenous Peoples. Indigenous Peoples' advocacy in scientific assessments is a major area of Indigenous human rights organizing

relating to science. Indigenous Peoples have interrupted the assumptions behind what knowledge counts as authoritative in scientific assessments, have taken steps to ensure that their own knowledge systems have significant or even leading roles, and have worked to ensure their rights are respected.

Indigenous Peoples often have advocacy strategies for enhancing their involvement. Consider the U.S. National Climate Assessment. In the 1990s, the National Aeronautics and Space Administration (NASA) and the Intertribal Council on Utility Policy sponsored a conference for Indigenous participants to address Indigenous perspectives and knowledge on climate change. A statement on Indigenous Knowledges stemming from the conference, the Albuquerque Declaration, was sent with the U.S. delegation to the IPCC COP on climate change that same year. The conference helped lead to the creation of a chapter, in the first National Climate Assessment, focused on Indigenous Peoples issues. The tradition of this chapter continued through the 2013, 2018, 2023, and the planned 2028 assessments, until the assessment was suspended in 2025. The majority of authors on this chapter are now scholars and scientists from Indigenous communities. The report features specialists in Indigenous Knowledge and Indigenous climate issues serving as authors across all chapters, and chapter-wide rules and guidelines for the use of Indigenous Knowledge.

Some multilateral assessments are just at the beginning phases of inclusion of Indigenous Peoples. For example, the International Coral Reef Initiative (ICRI), which pools together science and practice for conservation, has been attended by various Indigenous delegates over the years. But it was not until 2023 that the plenary included an opening workshop focused on Indigenous Knowledge and coral reefs. The opening workshop attempted to set a new tone for further actions, taken in later plenaries that would enhance Indigenous participation, especially given that coral reefs are among the aquatic habitats taken care of by a number of diverse Indigenous Peoples (Towle 2023).

IPBES has one of the most extensive processes for the inclusion of Indigenous Peoples through its policies on Indigenous and Local Knowledge. Following a decision by the third IPBES Plenary in 2015, members of Indigenous Peoples and Local Communities established the International Indigenous Forum on Biodiversity and Ecosystem Services (IIFBES), which is an open-ended, self-organizing network that aims to bring together members of Indigenous Peoples and local communities who are interested in the work of IPBES. The Network “aims to facilitate and enhance the effective participation of Indigenous Peoples and local communities in IPBES processes, taking into account the diversity of its members” (IPBES n.d.). Each IPBES assessment requires three (two if rapid) dialogues on Indigenous and Local Knowledge. There is an Indigenous and Local Knowledge Task Force, which plays key roles in helping to shape the processes of inclusion for each assessment under IPBES. At the last IPBES plenary in 2024, it was established that the next Global Assessment on Biodiversity will have an Indigenous and Local Knowledge focused chapter. IPBES has a well-defined process for how each assessment will solicit, review, and publish evidence and information from Indigenous and Local Knowledge sources. In 2025, IPBES recently held an international workshop on “enhancing methods for working with Indigenous and local knowledge in IPBES,” which convened a group of specialists that will issue a report with their recommendations on further changes that can be made to support Indigenous peoples’ participation in future IPBES work.

Indigenous Peoples are making larger strides in having higher levels of leadership in environmental science assessments. For example, one of the authors (Whyte) has served as a co-chair of the Climate and Biodiversity Assessment and as a U.S. Science Envoy (2023–2024), within which one of the goals was to support Indigenous leadership in multilateral science assessments. Indigenous scientists and scholars are showing up more often as having leading chapter author roles, including coordinating lead authors of chapters and formal reviewers of chapters. All in all, Indigenous advocacy in environmental science assessments has led to increased participation in various ways, including through mechanisms for participation, authorship and leadership roles, and rules and policies supporting Indigenous Knowledge.

31.5 Recommendations

We will conclude with two arguments mentioned earlier in this chapter. First, Indigenous Peoples whether as individual persons or as self-governing societies, ought to be equal participants and leaders in environmental science assessments sponsored by nations, both unilaterally and multilaterally. We will provide the reasons and examples of what interventions are most meaningful. Second, Indigenous Peoples should have the freedom and the support to organize and publish their own environmental science assessments using methods that arise from their communities' questions, knowledge systems, and social situations. We will give a few emerging examples of this practice of Indigenous-led environmental science assessments.

31.5.1 Equal participation and leadership

Indigenous Peoples have self-governance and stewardship responsibilities for taking care of entire territories. However, international bodies and nation states have denied Indigenous Peoples the right to participate as parties. For more than a century, the dominance of nations and their alliances have solidified to such a degree that it is not possible for the leaders of environmental science assessments to simply make Indigenous Peoples parties. In nation states with minimal to no consultative requirements, their specific agencies will design, plan for, and initiate assessments in the absence of early involvement and collaboration with Indigenous Peoples. Nation states will not open diplomatic relations to Indigenous Peoples at the earliest phases of signing treaties and agreements that create assessments, such as IPBES, IPCC, the ICRI, and others. As assessments are envisioned and organized, nations and secretariats assume that certain knowledge credentials, namely graduate degrees and professional experiences in non-Indigenous settings, are valued.

Equal leadership and participation would require a number of interventions, some of which are already being adopted in certain reports (Zurba et al. 2024). In terms of leadership, we mean the various roles played by people who are not serving in author roles in assessments. They are people at the highest levels who must approve the reports through their having a vote or a veto in a consensus process. Theoretically, this would mean that Indigenous Peoples would be parties or other governing members. However, in the absence of that in the short term, one solution is for nation states to include Indigenous Peoples as part of their diplomatic delegations. Providing Indigenous persons opportunities to serve in envoy and other diplomatic roles is critical. Nation states would have to establish processes for selection of Indigenous diplomats. There could also be established advisory, executive policies, and other mechanisms for influencing diplomatic teams. Some nation states have done this, including, for a brief time, the U.S.

A second intervention, at a high level, would involve ensuring that Indigenous persons are able to compete fairly for positions within the secretariats of national and multilateral processes that produce assessments. At one level, this would mean actual employment, for any position in the secretariat. At another level, if secretariats have national or other advisory committees, this would mean positions in the committees. The U.S. Global Change Research Program (USGCRP), for example, which oversees assessments such as the U.S. National Climate Assessment, established in 2016 a formal federal advisory committee that would assess how to sustain the report into the future. The committee had at least two specialists in Indigenous issues and at least one member of a Tribal Nation (Mosher 2016).

A third intervention regards the establishment of caucuses, task forces, and other fora for Indigenous Peoples to participate. One of the key issues, with any such intervention, is what level of access to decision-making and scientific processes do these bodies have. At one level, there are task forces that are staffed with elected members, whether by nation states or by relevant organizations, or through other processes. The IPBES Indigenous and Local Knowledge task force is selected through a nomination process. This legitimacy gives it influence. One of the barriers here is that a number of Indigenous

persons are not in the position to derive the type of support needed to participate. This is also the case for Indigenous authors in the IPCC processes. For example, university professors, government employees, and non-profit staff who are called to participate may receive time from their jobs, reimbursement of expenses, and chances to be rewarded in recognition for their service. While some Indigenous persons are in this position, a significant number are not. This is not so much an incentive issue, as it is a support issue. The types of background support needed to make participation of this kind possible is simply not immediately there for many Indigenous People. We only know of one environmental assessment—the Biodiversity and Climate Assessment—that has an Indigenous co-chair.

With regard to actual scientific processes, one of the key interventions is having Indigenous persons serve in author roles, including as chairs of entire assessments, chapter lead authors, chapter authors, or contributing or other author types. One of the key areas today for improvement is the support of Indigenous persons as authors. Often, Indigenous persons are not employed in occupations that allow for the volunteer service that scientists are able to perform. Many scientists work for universities, governments, and non-profit research institutions that allow them to carve out their employment time for service in assessment processes and get rewarded for it. Increasingly, authorship issues are coming to light as they are in critical need of solutions. A recent IPBES review of its work with Indigenous and Local Knowledge revealed some of the barriers that prevented Indigenous and local community members from having success as authors, in the few cases where they were in such roles (IPBES 2023).

These types of interventions would show how Indigenous rights, values, and knowledge can improve the quality and impact of scientific assessments, and what changes and transformations must be made within assessment processes. In the case of most scientific assessments, Indigenous advocacy has begun at the roots, establishing informal networks and forums to begin to express recommendations.

31.5.2 Indigenous-led assessments

Colonists, nations, businesses, and other actors, in different parts of the world, suppressed Indigenous Peoples' own institutions for knowledge, including research and educational institutions. Yet Indigenous Peoples, in various parts of the world, are reinstating their knowledge institutions. For example, in the U.S., there is the Tribal College and University system (Crazy Bull et al. 2020). Or in Indonesia, the Alliance of Indigenous Peoples of the Archipelago, *Aliansi Masyarakat Adat Nusantara* (AMAN), has established over 90 Indigenous schools that supplement students' public education.

Environmental science assessments are one of the major areas where Indigenous Peoples are calling for and beginning to establish their own activities. Different Indigenous-led groups have established multi-Indigenous assessments. In the experiences of the authors, it is extremely common for an Indigenous person, in a discussion about environmental science assessment, to call for the need for Indigenous-led assessments. Following the disastrous glacial lake outburst flood of 2024 in Khumbu (more commonly known as the Mount Everest region in Nepal), Sherpa community members from the region began their own climate change risk assessment process. A Local Adaptation Plans of Action (LAPA) was drafted in the summer of 2025 through community consultations in Khumbu for the Khumbu Pasanglhamu Rural Municipality as a step toward that direction. This document emphasized the interdependent ecosystem and the wellbeing of all beings as the guiding principle for climate change adaptation, drawing from Sherpa values of coexistence with all beings (humans and more-than-humans) of the land. A report on *Indigenous Peoples' Customary Governance for Sustainable Management of Natural Resources and Protection of Biodiversity* (Sherpa 2023) was produced by the CIPRED, an Indigenous research organization in Nepal, with largely Indigenous staff. It contextualizes the significance of customary institutions like the "Shagya" of the Tsum region in Nepal for biodiversity conservation.

He huringa āhuarangi, he huringa ao: a changing climate, a changing world, produced by a multidisciplinary Māori research team working across many research institutions, synthesizes the latest

climate change research through a Māori lens. The report identified the potential impacts, implications, mitigation, and adaptation strategies for whānau, hapū, iwi, and Māori business (Ngā Pae o te Māramatanga (NPM) and Manaaki Whenua—Landcare Research 2021). In the case of the Alliance of Indigenous Peoples of the Archipelago in Indonesia, they have innovated in further turning the tables. In one of their studies, spanning six communities, their own Indigenous and community-based research developed empirical information on six communities. Then they requested assessment of that information from an advisory body of non-Indigenous economists (AMAN Economy Team 2018).

In the U.S., one of the farthest along examples of an Indigenous-led assessment is the Status of Tribes and Climate Change Report, or STACC Report. The report is organized by the Institute for Tribal Environmental Professionals and receives some sponsorship from the Bureau of Indian Affairs in the U.S. The BIA is the agency charged with implementing the trust responsibility that the U.S. has with federally recognized Tribes. The STACC Report, volume 1, had over 90 authors, many of whom are Indigenous, and representing key roles in academia, Tribal Nations, and communities. The report featured both writing similar to the prose in science assessments, and testimonies that could directly report Indigenous Knowledge. The volume went through multiple layers of peer review and editing. The STACC report has recently published volume 2. While this volume is similar in terms of its process, there was a deliberate decision to have early-career Indigenous scholars and knowledge holders lead the process. This decision has led to innovation in chapter content (Status of Tribes and Climate Change Working Group (STACCCWG) 2021 and 2025).

Philanthropy, governments, multilateral organizations, and non-profits must consider the benefits, to Indigenous Peoples and beyond, of investing in independent Indigenous assessments and right authors. These are authors who are trusted by Indigenous Peoples and experts in ethics, experienced and knowledgeable on Indigenous environmental issues and rights, and capable of expressing knowledge synthesis in ways that are useful to the intended audiences. As with the STACC report, Indigenous-led assessments provide crucial confidence-building and training grounds for early-career scholars. Indigenous-led reports can also experiment with modes of expression that are novel, such as storytelling approaches, photography, video and animation, and different types of outputs intended to reach audiences that have different linguistic requirements and different expectations for how they are to receive critical information.

Note

- 1 'Āina Back Summit program, from which the quote is derived, is from the personal files of author Kyle Whyte.

References

- Akutagawa, Malia. 2019. "Nā Kai Po'olo'olo'u o Moloka'i: The Turbulent Seas of Moloka'i." In *Nourish: The Revitalization of Foodways in Hawai'i*, edited by Clare Gupta and Gillian Bostock Ewing. Extracurricular Press.
- Albuquerque, Ulysses Paulino, David Ludwig, Ivanila Soares Feitos, et al. 2021. "Integrating Traditional Ecological Knowledge into Academic Research at Local and Global Scales." *Regional Environmental Change* 21 (2): 45. <https://doi.org/10.1007/s10113-021-01774-2>.
- AMAN Economy Team. 2018. *Appraising Economic Performance of Indigenous Peoples' Sustainable Landscape Management*. AMAN, Climate and Land Use Alliance.
- Baptiste, Brigitte. 2020. "Virtual Expert Forum on SEEA EEA: The IPBES Values Assessment." Presentation. July 2020. IPBES, UNESCO, UNDP. Accessed September 9, 2025. https://seea.un.org/sites/seea.un.org/files/presentation_values_assessment_july_2020_un_seea.pdf.
- Belgian Earth Observation. 2021, August 9. "IPCC Published the First Part of Its 6th Assessment Report on Climate Change." Accessed February 26, 2025. <https://eo.belspo.be/en/news/ipcc-published-first-part-its-6th-assessment-report-climate-change>.
- Biddle, Justin B., and Anna Leuschner. 2015. "Climate Skepticism and the Manufacture of Doubt: Can Dissent in Science Be Epistemically Detrimental?" *European Journal for Philosophy of Science* 5 (3): 261–78. <https://doi.org/10.1007/s13194-014-0101-x>.

- Bohaker, Heidi. 2006. “‘Nindoodemag’: The Significance of Algonquian Kinship Networks in the Eastern Great Lakes Region, 1600–1701.” *The William and Mary Quarterly* 63 (1): 23–52. <https://doi.org/10.2307/3491724>.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Caretta, Martina Angela, and Ruth A. Morgan. 2021. “Special Issue on Indigenous Knowledge for Water-Related Climate Adaptation.” *Climate and Development* 13 (9): 761–5. <https://doi.org/10.1080/17565529.2021.1993627>.
- Carino, Joji. 2013. “Statement by Indigenous Peoples and Local Communities to the IPBES1 Plenary.” IPBES1 Plenary. Bonn, Germany: January 21–26, 2013. Forest Peoples Programme. Accessed September 9, 2025. <https://www.forestpeoples.org/fileadmin/uploads/fpp/migration/news/2013/02/Statement%20IPLCs%20to%20IPBES1%20Agenda%204%28b%29.pdf>.
- Carmona, Rosario, Graeme Reed, Stefan Thorsell, et al. 2023. “Analysing Engagement with Indigenous Peoples in the Intergovernmental Panel on Climate Change’s Sixth Assessment Report.” *Climate Action* 2 (1): 29. <https://doi.org/10.1038/s44168-023-00048-3>.
- Chesnais, Aude K., Joseph P. Brewer Kyle P., Whyte, Raffaele Sindoni Saposchnik, and Michael Kotutwa Johnson. 2024. “Reform Federal Policies to Enable Native American Regenerative Agriculture: Centering the Goals and Knowledge of Native Land Stewards in Federal Data and Definitions of ‘Climate Smart’ Agriculture Could Nourish Communities while Incentivizing Carbon Sequestration across Millions of Acres.” *Issues in Science & Technology* 40 (3): 50–54.
- Corbera, Esteve, Laura Calvet-Mir, Hannah Hughes, and Matthew Paterson. 2016. “Patterns of Authorship in the IPCC Working Group III Report.” *Nature Climate Change* 6 (1): 94–99. <https://doi.org/10.1038/nclimate2782>.
- Crazy Bull, Cheryl, Cynthia Lindquist, Raymond Burns, Laurel Vermillion, and Leander McDonald. 2020. “Tribal Colleges and Universities: Building Nations, Revitalizing Identity.” *Change: The Magazine of Higher Learning* 52 (1): 23–29. <https://doi.org/10.1080/00091383.2020.1693819>.
- David-Chavez, Dominique M., and Michael C. Gavin. 2018. “A Global Assessment of Indigenous Community Engagement in Climate Research.” *Environmental Research Letters* 13 (12): 123005. <https://doi.org/10.1088/1748-9326/aaf300>.
- Douglas, Heather. 2009. “Reintroducing Prediction to Explanation.” *Philosophy of Science* 76 (4): 444–63. <https://doi.org/10.1086/648111>.
- Elabbar, Ahmad. 2023. “The Curatorial View of Assessment and the Ethics of Scientific Advice: Beyond Decisional Autonomy towards Distributive Epistemic Justice.” *International Union for the History and Philosophy of Science and Technology (IUHPST) (Essay Prize in the History and Philosophy of Science)*. http://iuhpst.org/media/pdf/Elabbar_IUHPST_Prize_2023.pdf.
- Elabbar, Ahmad. 2025. “Justice as an Ideal for Climate Assessment.” PhD diss., Cambridge University. <https://doi.org/10.17863/CAM.116313>.
- Elliott, Kevin. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Fa, Julia E., James E.M. Watson, Ian Leiper, et al. 2020. “Importance of Indigenous Peoples’ Lands for the Conservation of Intact Forest Landscapes.” *Frontiers in Ecology and the Environment* 18 (3): 135–40. <https://doi.org/10.1002/fee.2148>.
- Ford, James. D., Will Vanderbilt, and Lea Berrang-Ford. 2012. “Authorship in IPCC AR5 and Its Implications for Content: Climate Change and Indigenous Populations in WGII.” *Climatic Change* 113: 201–13. <https://doi.org/10.1007/s10584-011-0350-z>.
- Ford, James D., Laura Cameron, Jennifer Rubis, et al. 2016. “Including Indigenous Knowledge and Experience in IPCC Assessment Reports.” *Nature Climate Change* 6 (4): 349–53. <https://doi.org/10.1038/nclimate2954>.
- Forest Peoples Programme. 2020. “Local Biodiversity Outlooks 2: The Contributions of Indigenous Peoples and Local Communities to the Implementation of the Strategic Plan for Biodiversity 2011–2020 and to Renewing Nature and Cultures: A Complement to the Fifth Edition of the Global Biodiversity Outlook.” Accessed September 9, 2025. <https://lbo2.localbiodiversityoutlooks.net>.
- Garnett, Stephen T., Neil D. Burgess, Julia E. Fa, et al. 2018. “A Spatial Overview of the Global Importance of Indigenous Lands for Conservation.” *Nature Sustainability* 1 (7): 369–74. <https://doi.org/10.1038/s41893-018-0100-6>.
- Goyal, Radhika. 2020, July 21. “Pasang Dolma Sherpa, Indigenous Peoples Representative to the U.N., Speaks With GlacierHub.” *State of the Planet*. <https://news.climate.columbia.edu/2020/07/21/pasang-dolma-sherpa-interview/>.
- Gundersen, Torbjørn. 2020. “Value-Free yet Policy-Relevant? The Normative Views of Climate Scientists and Their Bearing on Philosophy.” *Perspectives on Science* 28 (1): 89–118. https://doi.org/10.1162/posc_a_00334.

- ILO, International Labour Organization. 2020. "Implementing the ILO Indigenous and Tribal Peoples Convention No. 169: Towards an Inclusive, Sustainable and Just Future." <https://www.ilo.org/publications/implementing-ilo-indigenous-and-tribal-peoples-convention-no-169-towards>.
- IPBES. 2013. "The Contribution of Indigenous and Local Knowledge Systems to IPBES: Building Synergies with Science." June 9–11, Tokyo. Accessed September 9, 2025. <https://unesdoc.unesco.org/ark:/48223/pf0000225242>.
- IPBES. 2023, August 21. "Review of the Inclusion of Indigenous and Local Knowledge in IPBES Functions and Deliverables." Accessed September 9, 2025. <https://files.ipbes.net/ipbes-web-prod-public-files/2023-06/Review%20of%20Indigenous%20and%20local%20knowledge%20in%20IPBES.pdf>.
- IPBES. n.d. "About: Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services." Accessed December 14, 2025. <https://www.ipbes.net/about>.
- IPCC. n.d. "About: Intergovernmental Panel on Climate Change." Accessed September 9, 2025. <https://www.ipcc.ch/about/>.
- Jebeile, Julie. 2022. "Values and Objectivity in the Intergovernmental Panel on Climate Change." In *Questioning Experts and Expertise*, edited by Maria Baghramian and Carlo Martini. Routledge.
- John, Stephen. 2015. "The Example of the IPCC Does Not Vindicate the Value Free Ideal: A Reply to Gregor Betz." *European Journal for Philosophy of Science* 5 (1): 1–13. <https://doi.org/10.1007/s13194-014-0095-4>.
- John, Stephen. 2017. "From Social Values to P-Values: The Social Epistemology of the Intergovernmental Panel on Climate Change." *Journal of Applied Philosophy* 34 (2): 157–71. <https://doi.org/10.1111/japp.12178>.
- Kaltmeier, Olaf. 2016. "Mexican Indigenismo in a Hemispheric Context—Elements for a Historiography of Inter-American Entanglements in the First Half of the Twentieth Century." In *Mobile and Entangled America(s)*, edited by Maryemma Graham and Wilfried Rausser. Routledge.
- Klenk, Nicole, Anna Fiume, Katie Meehan, and Cerian Gibbes. 2017. "Local Knowledge in Climate Adaptation Research: Moving Knowledge Frameworks from Extraction to Co-Production." *WIREs Climate Change* 8 (5): e475. <https://doi.org/10.1002/wcc.475>.
- Lacey, Hugh. 1999. "Scientific Understanding and the Control of Nature." *Science & Education* 8: 13–35. <https://doi.org/10.1023/A:1008678413285>.
- Longino, Helen E. 1990. "Feminism and Philosophy of Science." *Journal of Social Philosophy* 21 (2–3): 150–9. <https://doi.org/10.1111/j.1467-9833.1990.tb00287.x>.
- Ludwig, David. 2016. "Overlapping Ontologies and Indigenous Knowledge. From Integration to Ontological Self-Determination." *Studies in History and Philosophy of Science Part A* 59: 36–45. <https://doi.org/10.1016/j.shpsa.2016.06.002>.
- Ludwig, David, and Charbel N. El-Hani. 2020. "Philosophy of Ethnobiology: Understanding Knowledge Integration and its Limitations." *Journal of Ethnobiology* 40 (1): 3–20. <https://doi.org/10.2993/0278-0771-40.1.3>.
- Ludwig, David, Charbel N. El-Hani, Fabio Gatti, et al. 2024. "Transdisciplinary Philosophy of Science: Meeting the Challenge of Indigenous Expertise." *Philosophy of Science* 91 (5): 1221–31. <https://doi.org/10.1017/psa.2023.127>.
- Massimi, Michela. 2025. "Local Knowledges and the Right to Participate in Science." *Philosophy of Science First View*: 1–25. <https://doi.org/10.1017/psa.2025.4>.
- McGregor, Deborah. 2020. "Indigenous Environmental Justice: Towards an Ethical and Sustainable Future." In *Routledge Handbook of Critical Indigenous Studies*, edited by Brendan Hokowhitu, Aileen Moreton-Robinson, Linda Tuhiwai-Smith, Chris Andersen, and Steve Larkin. Routledge.
- Mosher, Katie. 2016, June 29. "NOAA Establishes New Panel to Guide Sustained National Climate Assessment." *Sea Grant North Carolina*. Accessed February 26, 2025. <https://ncseagrant.ncsu.edu/noaa-establishes-new-panel-to-guide-sustained-national-climate-assessment/>.
- Ngā Pae o te Māramatanga (NPM) and Manaaki Whenua—Landcare Research. 2021, October 21. "He huringa āhuarangi, he huringa ao: A Changing Climate, a Changing World." Accessed February 22, 2025. <https://www.landcareresearch.co.nz/news/release-of-te-ao-maori-climate-change-report>.
- PACT (Pathways Alliance for Change and Transformation). 2023. "Indigenous Peoples, Local Communities, and Trade in Terrestrial Carbon Markets: Time to Recognize Their Rights, Support their Organizations, and Impose a Moratorium on Terrestrial Carbon Trade." Pathways Alliance for Change and Transformation.
- Petersen, Arthur. 2021. "Values and Accountability in Science Advice: The Case of the IPCC." In *Science, Values, and Democracy: The 2016 Descartes Lectures*, edited by Ted Richards. Consortium for Science, Policy & Outcomes.
- Ramirez, Renya K. 2009. "Henry Roe Cloud: A Granddaughter's Native Feminist Biographical Account." *Wicazo Sa Review* 24 (2): 77–103. <https://www.jstor.org/stable/40587782>.
- Ressorio C., Giulia De Fusco, David Ludwig, Charbel N. El-Hani, and Esther Turnhout. 2025. "Caring Policy-Relevant Knowledge? The Case of the Brazilian Platform for Biodiversity and Ecosystem Services." *Environmental Science & Policy* 171: 104170. <https://doi.org/10.1016/j.envsci.2025.104170>.

- Rights and Resources Initiative, Woodwell Climate Research Center, and Rainforest Foundation US. 2021. *Significance of Community-Held Territories in 24 Countries to Global Climate*.
- Sherpa, Pasang Dolma. 2019, December 30. "The Historical Journey of Indigenous Peoples in Climate Change Negotiation." *IUCN*. Accessed February 13, 2025. <https://iucn.org/news/commission-environmental-economic-and-social-policy/201912/historical-journey-indigenous-peoples-climate-change-negotiation>.
- Sherpa, Pasang Dolma. 2023. *Indigenous Peoples' Customary Governance for Sustainable Management of Natural Resources and Protection of Biodiversity*. CIPRED.
- Sherpa, Pasang Yangjee. 2025. "Relevance of the Sixth IPCC Assessment Report to Indigenous Lived Realities." *AlterNative: An International Journal of Indigenous Peoples*. <https://doi.org/10.1177/11771801251362621>.
- Smith, Heather A., and Karyn Sharp. 2012. "Indigenous Climate Knowledges." *Wiley Interdisciplinary Reviews: Climate Change* 3 (5): 467–76. <https://doi.org/10.1002/wcc.185>.
- Status of Tribes and Climate Change Working Group (STACCCWG). 2021. "Status of Tribes and Climate Change Report." Institute for Tribal Environmental Professionals, Northern Arizona University.
- Status of Tribes and Climate Change Working Group. 2025. *Status of Tribes and Climate Change*, edited by B. M. Panek, Vol. 2. Institute for Tribal Environmental Professionals, Northern Arizona University. <http://nau.edu/staccreport>
- Tengö, Maria, Eduardo S. Brondizio, Thomas Elmqvist, Pernilla Malmer, and Marja Spiereburg. 2014. "Connecting Diverse Knowledge Systems for Enhanced Ecosystem Governance: The Multiple Evidence Base Approach." *Ambio* 43 (5): 579–91. <https://doi.org/10.1007/s13280-014-0501-3>.
- Tengö, Maria, Rosemary Hill, Pernilla Malmer, et al. 2017. "Weaving Knowledge Systems in IPBES, CBD and Beyond—Lessons Learned for Sustainability." *Current Opinion in Environmental Sustainability* 26–27: 17–25. <https://doi.org/10.1016/j.cosust.2016.12.005>.
- ThinkLandscape. n.d. "Pasang Dolma Sherpa." Accessed February 13, 2025 <https://thinklandscape.globallandscapesforum.org/collection/pasang-dolma-sherpa/>.
- Towle, Erika K. 2023. "International Meeting Brings Global Leaders Together to Address Coral Reef Conservation." *NOAA Coral Reef Conservation Program*. <https://www.coralreef.noaa.gov/aboutcrp/news/featured-stories/dec23/ICRI2023.html>.
- Tunks, Andrea. 1997. "Tangata Whenua Ethics and Climate Change." *New Zealand Journal of Environmental Law*, 1, 67–124.
- UNESCO, IPBES, and United Nations University. 2013. "The Contribution of Indigenous and Local Knowledge Systems to IPBES: Building Synergies with Science." IPBES Expert Meeting Report. June 9–11, 2013. <https://unesdoc.unesco.org/ark:/48223/pf0000225242>.
- United Nations (UN). 1982. "Convention Providing for Creation of the Inter-American Indian Institute (with Annex)." Concluded at Mexico City on November 29, 1940.
- van Bavel, Bianca, Joanna Petrasek MacDonald, and Dalee Sambo Dorough. 2022. "Indigenous Knowledge Systems." In *A Critical Assessment of the Intergovernmental Panel on Climate Change*, edited by Kari De Pryck and Mike Hulme. Cambridge University Press. <https://doi.org/10.1017/9781009082099.017>.
- Whyte, K. (2018). "What Do Indigenous Knowledges Do for Indigenous Peoples." *Traditional Ecological Knowledge: Learning from Indigenous Practices for Environmental Sustainability*, 57–82.
- WWF, UNEP-WCMC, SGP/ICCA-GSI, et al. 2021. "The State of Indigenous Peoples' and Local Communities' Lands and Territories."
- Zurba, Melanie, Sandie Suchet-Pearson, Ryan Bullock, et al. 2024. "Enhancing Meaningful Indigenous Leadership and Collaboration in International Environmental Governance Forums." *Environmental Science & Policy* 162: 103864. <https://doi.org/10.1016/j.envsci.2024.103864>.

CONTROVERSIES IN VACCINE POLICY

Maya J. Goldenberg

32.1 Introduction: “We need to follow the science”

A February 7, 2022 sitting of the Canadian legislature was focused on the occupation taking place outside the doors of the Parliament building by the so-called Freedom Convoy, a populist protest first initiated to object to COVID vaccine mandates required for Canadian truck drivers crossing the US border, and later evolved to protest most COVID-related public health restrictions and requirements.¹ At this meeting of the House of Commons, Conservative MP and leader of the Official Opposition Candice Bergen gestured to the audible noise of blaring air horns² from the big rig trucks blocking roads throughout downtown Ottawa and asked the governing Liberal Party when Canadians could expect COVID vaccine mandates and public health restrictions to be removed. The response from Government House Leader Mark Holland was, “we need to follow science, using evidence not politics” ([Global News 2022](#)). But “following the science” masks value judgments rather than avoiding them.

Policy controversies over vaccines and other hot-button science policy issues (e.g., climate change) illustrate the inadequacy of positioning science as the “trump card” over values. These controversies arise from value disagreements that are insufficiently addressed, and science is often instrumentalized for shutting down debate, dialogue, and other means for addressing social and political disagreement. It was common during the COVID pandemic to hear reassurances from political leadership in many nations that they were “following the science.” This was rhetoric that said nothing about their political processes—Were they democratic and consultary?—or their policies—Were they reasonable, measured, and effective? While science can tell us if vaccines are safe and effective, for example, it cannot tell us if it is morally justified to compel unwilling members of the public to get vaccinated. Critics raised those moral and political concerns, often with hostility, as the siege on downtown Ottawa and many threats of violence directed at scientific and political leadership demonstrated ([Nogrady 2021](#)). Those complaints were maligned as misinformed, and the complainants were cast as racists, radicalized misanthropes, and unmoored agitators. Canadian Prime Minister Justin Trudeau described the Canadian trucker convoy as hateful, abusive, and racist (and indeed some events of the Ottawa siege supported this description), and the objections to vaccine mandates as showing “disrespect to science” ([Naughtie 2022](#)). Public attention was thereby directed farther away from policy disagreements over vaccines.

This chapter steps back from the populist framing of political opposition between “the virtuous” and “the vile” ([Friedman 2017](#)) to challenge the ill-founded positioning of science vs. politics and to

explicate the value disagreements that arise in vaccine politics and policy. The goals of this analysis are both conceptual clarity and practical: Instrumentalizing science, ignoring value conflicts, and dehumanizing opposition makes things worse. Instead, vaccine policy controversy requires more charitable understanding of opposing views and, with that understanding, deliberative recourse through political channels. This is difficult work in this political moment, with the recent conclusion of global pandemic³ in which vaccine mandates were a lightning rod for numerous social and political discontents and disagreements. But the work is unavoidable.

This chapter will address the values associated with vaccination (Section 32.2), followed by the values specific to vaccine policies, especially mandatory vaccination policies (Section 32.3). Attention is paid to a shift in the primary political goals associated with contemporary vaccine mandates, from their introduction (starting in the 1960s) as school-entry requirements aimed at building social norms of vaccination, to a punitive and/or coercive political instrument starting in the 2010s. With that shift in policy frame came a correlating change in vaccination protest speech from safety concerns regarding vaccination to “medical freedom.” The contemporary political framing of vaccine policy controversy as *science vs. freedom* is lamentable (Section 32.4). My argument, which is reminiscent of criticisms of evidence-based policy since the early 2000s, maintains the impracticality of scientizing politics when competing values are in conflict and insists on the inescapability of democratic deliberation.

32.2 Vaccines and values

When vaccine hesitators and refusers make statements like “vaccines are not safe,” “vaccines are not effective,” and “vaccines are not necessary” that belie expert opinion and scientific consensus claiming otherwise, one might reasonably conclude that there is some misunderstanding or misinformation at root. Indeed, this is the dominant causal interpretation of vaccine hesitancy.

Health outreach directed at the publics build on this assumption and follow a communications tactic of addressing the perceived “knowledge deficit” by identifying the mistaken or missing information and attempting to provide the corrective “facts” about vaccines (Goldenberg 2016). Government communications to Canadians during the COVID pandemic, for example, encouraged the publics to “Get the facts. Visit [Canada.ca/COVID-vaccine](https://www.canada.ca/covid-vaccine) to learn more.” The WHO offered COVID advice to the publics using a “mythbusters” format of infographics and videos that debunked misinformation and presented the facts based on up-to-date clinical research (WHO 2022).

While there is no harm in sharing and amplifying reliable information regarding public health and vaccines, it remains a questionable choice as the frame and focus of the great majority of public health communications. This is because misinformation and misunderstanding are unlikely to be a significant cause of vaccine hesitancy. Once again, it is values and not scientific facts that need attention.

If misunderstanding and misinterpretation of the scientific evidence were the problems, the provision of accurate information and education would presumably correct this. Many years of education campaigns regarding childhood vaccination have not substantially decreased numbers of vaccine refusers and vaccine hesitancy has tipped upward (Goldenberg 2016). Furthermore, education levels are not reliable predictors of people’s acceptance or refusal of vaccines.

Public mistrust of scientific institutions better explain vaccine hesitancy than does scientific misunderstanding and/or misinformation (Goldenberg 2021). Mistrust primes individuals to seek out and accept alternative claims by dubious experts regarding vaccines (rather than the converse assumption that misinformation causes mistrust) (Williams 2023). There is ample social science research examining parents’ vaccine hesitancy and refusal; survey respondents and members of focus group participants frequently cite discomfort with medicine’s close ties to pharmaceutical and biologics

industries (e.g., [EKOS Research Associates 2018](#)). Racialized communities hit hardest by COVID morbidity and mortality were polled to have higher than average levels of COVID vaccine hesitancy both before the vaccine was approved and in the early months of the vaccine roll-out. Testimonials revealed that community members do not trust the vaccines, the health care system, and/or the government. To illustrate, see the 2020 *New York Times* investigation “I Won’t Be Used as a Guinea Pig for White People” ([Hoffman 2020](#)).

Those misgivings are far from the wild conspiracy theories that researchers have tried to link to “anti-vax” views. Instead, they are fissure points that are then seemingly confirmed when parents’ questions about childhood vaccines are shut down by their health care providers ([Goldenberg 2021](#), 157–8). Or they build upon histories and contemporary experiences of medical racism, structural inequalities, and social injustices that are, in fact, what led to racialized communities being hit so hard by COVID (e.g., [Mosby and Swidrovich 2021](#)).

Yet many vaccine criticisms *sound* like scientific misunderstanding when the critics, most of whom are non-experts, cherry-pick dubious studies and put so much energy into arguing technical points about adjuvants, boosters, titers, spike proteins, and mRNA platforms. While arguments over vaccines are often centered on the science—with vaccine advocates pointing to the strong consensus on vaccines and vaccine skeptics curating research to generate narratives of suppressed science showing vaccines to be unsafe, ineffective, or unnecessary—the science largely serves as placeholders for the values at stake. At issue in vaccine controversy, much like heated climate change disagreements ([Sarewitz 2004](#)), is what follows practically from accepting the science as true ([Goldenberg 2021](#)). The scientific evidence serves as proxy for the values that are on the line, such as individual liberties vs. common goods, medical progress vs. “natural” living, what duties we have toward others and toward future generations, among other value debates. None of these issues are easily settled and, importantly, none will be settled by the science of vaccines.

Vaccine debates, then, are about much more than vaccines. Instead, they capture a host of temporally, geographically, and historically specific concerns. In liberal democratic societies, those concerns include how technology shapes our lives, who decides on and regulates those technological intrusions, the relationship of knowledge and power (i.e., whether science works for the people or for corporate interests), government overreach and individual liberty, community cohesion, health disparities, income inequality, globalization, and more.

Further, because vaccination requires government-led coordination, funding, and enforcement to achieve the collective goal of public health, vaccine choices are unavoidably tied to politics and policy. This brings us to the focus of the chapter, namely controversies in vaccine policy. In addition to imbricating the ethical tension between individual choice and collective need, vaccine policies often serve as de facto gauges of public dis/approval of their government. Government agencies invest in purchasing and access to vaccines (e.g., through cost and availability) and engage in health communications to advance public health objectives. The publics thereby interpret the act of vaccination to be not only a health choice but a political act. In jurisdictions where COVID vaccines were accessible to the populace and encouraged and/or mandated as a collective effort to end the pandemic, vaccination affirmed trust in government and the institutional structures managing the public health crisis, while refusal repudiated public health infrastructure and government leadership.

32.3 Vaccine policies

The success of vaccination campaigns hinge on public trust in government institutions ([Adhikari, Yeong Cheah, and von Seidlein 2022](#); [Goldenberg 2021](#); [Krastev et al. 2023](#)) because so much legislative decision-making and oversight go into these public health efforts (see, for example, “Regulating vaccines for human use in Canada” [[Government of Canada 2020](#)]). Governments make numerous

decisions about whether and how to promote individual vaccines (Navin and Attwell 2019). First is the determination of whether to license a vaccine for sale and administration within its jurisdiction. Second is which of its licensed vaccines will be recommended, to whom, and how (e.g., through public outreach and/or by being included in a recommended vaccine schedule). Third, a government must determine whether to publicly fund, or facilitate the funding of, individual vaccines in order to improve access. Finally, a government must decide whether the vaccine should be mandated, and what will be the magnitude of the burdens imposed by the state on those who do not comply (Navin and Attwell 2019). Government oversight does not end once policy has been put in place. Ongoing safety and surveillance monitoring necessitate that policies may change, for instance, when population-level disease frequencies increase or decrease, vaccine safety concerns arise or subside, and the public respond favorably or unfavorably toward the health communications and/or vaccination requirements placed on them.

The remainder of this section on vaccine policies will focus on mandatory vaccine policies (“vaccine mandates”), the most controversial vaccine policy option, and the clearest means for illustrating how values infuse vaccine policy determination, as policies are political instruments meant to affirm or attain explicitly political goals.

32.3.1 Vaccine mandates

The majority of the public, professional, and scholarly attention to vaccine policy is focused on vaccine mandates, a variety of policy options for increasing vaccination rates by requiring target populations to receive the vaccine. Vaccine mandates vary in terms of scope, sanctions, and severity (Navin and Attwell 2019). Despite infringing on personal choice, arguments from a variety of rights-based ethical and political traditions converge on the permissibility of careful and measured deployment of state-sanctioned vaccine mandates. There are also consequentialist, justice, and virtue ethics arguments supportive of mandatory vaccination under the right conditions (e.g., Boyneburgk and Bellazzi 2022; Reiss and Caplan 2020).

Whereas liberal states normally uphold personal choice regarding health care interventions, exceptions are made for emergency conditions (e.g., epidemics, war), and situations that threaten the legal, economic, or political foundations of individual rights themselves (Matthews and Menzel 2023, 719–20). Liberal political frameworks, then, include conditions under which the state may override its *prima facie* commitments to individual rights and liberties, including overriding consent when the individual or institutional actions pose a reasonably demonstrable threat to others. Vaccine jurisprudence follows this course. School-entry vaccine mandates have survived legal challenges (for an overview of US court challenges, see Hodge and Gostin 2002), as have numerous workplace vaccine mandates. During COVID, for example, Canadian labor arbitrations upheld many workplace COVID vaccine mandates, even those with severe penalties like possible termination of employment upon refusal, finding vaccine requirements to be reasonable given the gravity of a global pandemic and the right to a safe workplace in contexts where accommodations for unvaccinated workers (e.g., working remotely) could not be implemented (Hudson 2022).⁴ As well, legal opinion did not support the claim, made by numerous vaccine mandate critics, that COVID vaccine requirements violate Canadian Charter of Rights and Freedoms (Bogart 2021).⁵

The general consensus, then, is that vaccine mandates are morally and legally justified when utilized prudently, with attention to the proportionality of the public health threat relative to the individual and collective benefit offered by a safe and effective vaccine. The harm of infringement of individual liberty should be minimized by implementing the least restrictive policy among the available alternatives. Mandatory vaccination “should only be used when all other reasonably available options have been exhausted. Free informed choice to participate is the most desirable option,

followed by a variety of nudging strategies aimed at promoting higher rates of immunization” (Cave 2017). Vaccines must also be easily accessible to all targeted demographics before mandates are considered (Omer, Betsch, and Leask 2019). Mandates enter into policy consideration because education, persuasion, and nudging are often insufficient to achieve community protection (also known as, and unfortunately termed, “herd immunity”). Although scholars may disagree about the thresholds for implementation, they agree that there are circumstances under which mandatory vaccinations are permissible and even necessary given the overriding public health importance of vaccination (Matthews and Menzel 2023, 721).

32.3.2 Vaccine mandates to establish social norms

While the ethics of vaccine mandates literature tends to focus on the moral permissibility of coercive interventions, vaccine mandates have not always served the purpose of coercion. Attwell and Navin (2023) detail the history of school-entry vaccine mandates in the US, describing these policies as the “lynchpin of American vaccine governance since the mid-20th century.” When state-level school-entry mandates were first introduced in the 1960s, they were not clearly at odds with American predilection for individual liberties because the new mandates did not aim to coerce. There was bipartisan political support for policy efforts to create social norms of vaccination among parents of school-aged children, who were “basically willing” but “needed to be socialized into giving their children the newly available vaccines” (Attwell and Navin 2023, 672). Mandates would boost compliance and thereby help establish that norm (Attwell and Navin 2023, 672; Omer, Betsch, and Leask 2019). School-entry vaccine mandates helped stave off measles outbreaks in schools by increasing coverage, and they importantly included non-medical exemptions (NMEs) in all but two US states, thereby stopping shy of governing parents’ vaccine behavior.

Scholars examining other liberal democracies than the US have similarly pointed to the normalizing of vaccination enabled by introducing vaccine mandates that include non-coercive costs and benefits favoring vaccination.⁶ Implementing a vaccine mandate conveys to the public that vaccination is not a personal choice, but rather something expected as a member of a population, similar to taxes (Giubilini 2020). Mandates also formalize a moral duty owed to protect others from harm (Navin 2013; Pierik 2018). In France, one rationale given for expanding school-entry vaccine requirements from three to eleven vaccines in 2018 was to message that the previously elective pediatric vaccines were important and necessary for protecting children’s health (Ward, Colgrove, and Verger 2018). Mandates thereby need not be understood to be inherently coercive, intentionally punitive, and thereby predictably damaging to public trust (Goldenberg et al. 2023). By creating social norms around vaccination, making vaccine refusal more costly or inconvenient than vaccination, and by providing assurance that others are also making their contribution, mandates can improve vaccine uptake rates with minimal public antagonism. Omer, Betsch, and Leask (2019) recommend policymakers to “mandate vaccination with care” to strike the right balance of prescription without antagonism.

32.3.3 Vaccine mandates as coercive policies

An increase in measles outbreaks in high-income countries in the 2010s (Poland and Jacobson 2012) put a perceived weak point in mandatory vaccination policy under critical scrutiny: The NME (e.g., Attwell and Navin 2023; Bednarczyk et al. 2019; Navin and Largent 2017; Olive et al. 2018).

In the US, for example, measles infection peaked in the late 2010s and the country nearly lost its measles elimination status in 2019. The NME, available in all but two US states’ school-entry vaccine policies, was identified as the problem in otherwise good policies. Indeed, use of the NME by vaccine-hesitant and vaccine-refusing parents had increased over the past two decades

and the connection between under-vaccinated pediatric populations and measles outbreaks was already well established (Bednarczyk et al. 2019). Additionally, US states with more cumbersome or restrictive NME access had lower infection rates (Omer et al. 2012). Legislators capitalized on public frustration toward the so-called anti-vaxxers and used the political moment to introduce severe restrictions and difficulty for obtaining NMEs. One state, California, boldly eliminated all NMEs for school-entry vaccination in 2016 (NB: Medical exemptions remained untouched). This new policy, and others like it that followed in other states and countries, signified a turn in vaccine policy toward coercion.

While adjustments to NMEs seem warranted and prudent in the context of increased measles cases and increased use of an exemption that was designed to be administered sparingly (so as to keep community-level protective immunity secure), the policy changes were perceived as punitive and “hard-line” (Rainford and Greenberg 2015). Indeed, elimination of NMEs radically reshaped vaccination governance by undermining previous “soft” policy pressures (e.g., the use of behavioral strategies like nudging) and instead constrained parents’ vaccination behavior by removing choice. These new, more restrictive policies were also criticized for being “quick fixes” to suboptimal vaccination rates that clumsily antagonized the publics rather than attending to practical and sociological barriers to vaccination, namely poor access to vaccines in underserved communities (Navin and Attwell 2019; Omer et al. 2019) and declining public trust in scientific and health governance (Goldenberg 2021).

Attwell and Navin (2023) characterize the legislative attention to delimiting NMEs in pediatric vaccine policy throughout the latter half of the 2010s as a regime change in vaccine policy from the earlier “mandate & exemption” period beginning in the 1960s, described above as an effort to normalize vaccination through gentle (i.e., non-coercive) policy, to a regime of “unencumbered mandates.” Without exemptions in place to satisfy those parents unwilling to vaccinate their children, vaccine mandates were repurposed as coercive policy. Numerous difficulties followed: More galvanized and better organized vaccine refusers, more requests for medical exemptions, more physicians censured for providing those medical exemptions, more pressures on schools to monitor and enforce vaccine policies, and more political polarization overall. This heightening of vaccine-related politicization arguably arose because these long-standing school-entry mandates were never designed to discipline and punish wayward parents as a mode of vaccine governance (Attwell and Navin 2023, 672).

Around the same time that NMEs were under legislative scrutiny, the discourse of vaccine skepticism and refusal shifted from safety concerns to the language of *medical freedom*. Historically and to the present day, vaccine-critical rhetoric has rested on two principal claims: First, that vaccination is a dangerous procedure with risks that outweigh benefits, and second, that efforts to pressure or compel people to be vaccinated (or to vaccinate their children) violate individual rights (Schwartz 2012). A 2019 study of Facebook posts found that in recent years, arguments related to individual liberty had grown more prominent in anti-vaccination messaging, with vaccine refusers increasingly framing their choice as a civil right (Jamison et al. 2020). This shift leans on a history of American health libertarianism (otherwise known as the “health freedom movement”) that dates back to the 1800s (Grossman 2021, 10–23). It also served as a poignant counterresponse to the political climate and policy levers pushing toward more restrictive vaccination choice. This change in political speech by vaccine opponents starting 2014–2015 (recounted by Haelle 2015 and discussed in Goldenberg 2022), and later used as a rallying cry against all COVID pandemic public health measures, including vaccine mandates, spurred a rhetorical redirection by vaccine proponents as well. With the language of “freedom” usurped by vaccine skeptics and public health refusers, liberal governments and public health agencies became the unwitting enemies of personal choice, family autonomy, and individual liberty. In response to this unfortunate framing, vaccine advocates doubled down on the science supporting vaccination. What followed was a new vaccine policy discourse that pitted freedom against science.

32.4 Controversies in vaccine policy: Science vs. freedom

This brings us back to the Canadian House of Commons, where the governing Liberal Party promised to use science to steer national COVID vaccine policy and to (inexplicably) lead the country out of a severe public safety threat created by the “Freedom Convoy” siege on Ottawa. Vaccine policy controversy is now commonly framed as “science vs. freedom,” an uneven pitting of allegedly value-free truths against a fundamental value. This frame ought to be dismantled in favor of attending to the values at stake on both sides of vaccine mandate controversy. Vaccine mandate dissenters are clear that they value individual choice and find vaccine mandates as well as other public health collective measures to be unwanted infringements on that value. On the other side, “follow the science” is less clear in its value orientation by allegedly rising above the political fracas in favor of facts and truth. However, the political strategy of “follow the science” is a value-laden strategy insofar as it prioritizes the common good (public health) over individual liberty, as well as privileging technocratic governance, that is, decision-making by a selective group of experts rather than more democratic methods that involve public consultation. It also quiets dissent as “misinformed” or worse. This appeal to science sidesteps political discourse and tolerates the unintended but not unexpected consequence of societal polarization on ideological grounds. “Follow the science” strategies, much like the aspirational “evidence-based policymaking” movement that began at the turn of this century (Hammersley 2005; Marston and Watts 2003), (ironically) politicize science by insisting that science can stand outside of politics.

Technocratic governance is presumably not the political goal of all vaccine mandate supporters. The millions of global citizens who not only chose COVID vaccination but supported mandates were arguably yearning for the aspirations of evidence-based/science-led politics, including health and safety, and a calm and civil political climate. Vaccine mandates do not, of course, cultivate such a climate (nor does outrage directed at those who refuse vaccination). Vaccine mandates may underperform, as widespread vaccination and disease eradication do not easily arise from restrictive policies introduced into hostile political climates.

As per the previous discussion of “mandate with care,” vaccine mandates that are insensitive to context will be perceived by the publics as punitive. COVID vaccine mandates were largely seen this way due to design and communication. In terms of design, many vaccine mandates were criticized for being applied to low-risk populations (e.g., school-aged children), for ignoring natural immunity from prior COVID illness, for providing very selective medical exemptions, and for not providing NMEs. The penalties were perceived to be severe: Restricting access to public spaces, education, and employment—with insufficient regard for the socioeconomic, political, and cultural harms created by constrained choice, as well as the inequitable distribution of these harms among different social groups. COVID vaccine mandates did not strike the delicate balance between increasing vaccination without promoting public backlash. Communications from global leaders suggested that this balancing act was not a priority.

French President Emmanuel Macron said in a January 2022 interview with *Le Parisien* that his government’s proposed “vaccine passport,” which limits access to public spaces to vaccinated people only, was meant to “piss off” the unvaccinated (he used the vulgar term “emmerder” in French) (Beaumont et al. 2022). In that same interview, he made the offensive claim that somebody who acts irresponsibly is no longer a citizen (Ataman, Mawad, and Rob Picheta 2021). Other world leaders were equally derisive. Former British Prime Minister Tony Blair described unvaccinated people as “not just irresponsible. You’re an idiot” (Davidson 2021), then-premier of Western Australia, Mark McGowan preferred the descriptors “crazy” and “deranged” (Fiore 2021), while then-Romanian Prime Minister Florin Cîtu likened vaccine protesters to “terrorists” (Xinhua 2021). These are only a few examples.

This type of “othering” those with whom we have political disagreements, whereby their rationality and characters are dismissed and their arguments derided as “misinformation,” might lead

one to believe that deliberation is impossible. We have a handful of highly used tropes to sediment this view—“post-fact society,” and “truth decay” brought on by “information warfare” (see [Hwang 2020](#))—and to seed the idea that “follow the science” might be all that is left to do.

Yet it has already been mentioned that “following the science” as a response to policy questions is evasive at best, illiberal at worse, and prone to public backlash. The hard work of participatory democratic politics is not evaded; instead, its importance is underscored. When the science is mostly not settled, and it is multiple (economic, immunological, sociological, virological, behavioral, and more), which science is being followed? How are the trade-offs between the plurality of diverse war-rants for different courses of action being handled?

Furthermore, policymaking does not follow a linear trajectory from science to action. This argument has been amply made by critics of “evidence-based policymaking” (EBPM), an aspirational concept first introduced to political circles in the late 1990s/early 2000s (e.g., [BBC News 2010](#)). The goal was to free policymaking from the everyday pressures of politics by getting the facts right. In *The Politics of Evidence-Based Policymaking*, political scientist Paul Cairney argued that EBPM so conceived is unrealistic. Cairney implores readers to drop the naive understanding and “ill-considered aspirations” of EBPM as “comprehensive rationality” whereby “policymakers are able to generate a clear sense of their preferences, gather and understand all relevant information, and make choices based on information” (2016, 2–5). Policymaking is better understood as “bounded rationality” where “policymakers have unclear aims, limited information, and unclear choices” and use various shortcuts to reach policy decisions ([Cairney 2016](#), 5). The policy process is defined by

how problems are ‘framed’ by their advocates and how they are understood by the policymakers held responsible for solving them. It is about the power to ignore or pay attention to particular studies, to link the evidence of a policy problem to a particular solution, and to ensure that policymakers have the motive and opportunity to turn a solution into policy.

(Cairney 2016, 6)

This account highlights how policymaking needs more scrutiny in order to make the process more inclusive and comprehensive, and so the emphasis should be on political legitimacy rather than more rhetoric of following the science. Policymaking, by this account, needs an expansion rather than contraction of the sphere of policy influence.

To bring this discussion back to the parliamentarians in Ottawa, science cannot lead politics and policy. We have to decide what we care about, and what our goals are. These are political questions that require political deliberation, unavoidably so. Only then can science inform and enable the direction in which the demos have chosen to go. Those are decisions of value for which the scientific experts have no special authority to decide, and democratic deliberation, difficult as it may be, is the legitimate means.

32.5 Conclusion

The argument in this chapter has been that vaccine policies have come to signify many political hopes and desires—for health and safety, civil discourse, political calm, and rational agreement—and pronounced political fears concerning government overreach and infringement of rights and liberties. The framing of the conflict as “science vs. freedom” denies the values underlying the “follow the science” stance. In response, “medical freedom” arguments misrepresent liberty in absolutist terms. Instead, individual freedom and vaccination requirements have a long moral and legal history of concurrence, and attention should be focused on establishing ethical vaccine requirements rather than denying the possibility. Both sides play into political hyperbole that disable the possibility of political agreement.

Notes

- 1 Both Canada and the US had exempted unvaccinated cross-border truckers from COVID-19 vaccine requirements to prevent exacerbating existing supply chain disruptions, but that exemption in Canada ended on January 15, 2022, and the US exemption ended on January 22, 2022. Of the 120,000 Canadian licensed truck drivers who regularly served cross-border routes, approximately 85 percent were vaccinated against COVID-19 by January, leaving up to 16,000 Canadian truckers potentially affected by the new restriction that spurred the Freedom Convoy protest.
- 2 See [Crawford \(2022\)](#).
- 3 The World Health Organization (WHO) declared COVID-19 to no longer be a global health threat and therefore removed the designation of “pandemic” in May 2023.
- 4 The rights of employers to collect data from employees about their COVID vaccination status was also upheld by the [Office of the Privacy Commissioner of Canada \(2023\)](#).
- 5 The Canadian Charter of Rights and Freedoms (the “Charter”) is a bill of rights entrenched in the Constitution of Canada.
- 6 Mandatory vaccination exists in some form in most countries, most commonly for childhood vaccine-preventable diseases ([Gravagna et al. 2020](#)).

References

- Adhikari, Bipin, Phaik Yeong Cheah, and Lorenz von Seidlein. 2022. “Trust Is the Common Denominator for COVID-19 Vaccine Acceptance: A Literature Review.” *Vaccine: X* 12: 100213. <https://doi.org/10.1016/j.jvaxc.2022.100213>.
- Ataman, Joseph, Dalai Mawad, and R. Rob Picheta. 2021. “Macron Says He Wants to ‘Piss Off’ the Unvaccinated.” *CNN*, January 5. <https://www.cnn.com/2022/01/05/europe/macron-unvaccinated-comments-vaccine-pass-intl/index.html>.
- Attwell, Katie, and Mark C. Navin. 2023. “Eliminating Nonmedical Exemptions: A Radical Shift in How Childhood Vaccine Mandates Govern.” *Expert Review of Vaccines* 22 (1): 671–80. <https://doi.org/10.1080/14760584.2023.2232446>.
- BBC News. 2010. What Matters Is What Works.” *BBC News*, June 4. http://news.bbc.co.uk/2/hi/programmes/more_or_less/8720866.stm.
- Beaumont, Olivier, David Doukhan, Pauline Théveniaud, Henri Vernet, and Marcelo Wesfreid. 2022. “Europe, Vaccination, Présidentielle... Emmanuel Macron Se Livre à Nos Lecteurs.” *Le Parisien*, January 4. <https://www.leparisien.fr/politique/europe-vaccination-presidentielle-emmanuel-macron-se-livre-a-nos-lecteurs-04-01-2022-2KVQ3ESNSREABMTDWR25OMGWEA.php>.
- Bednarczyk, Robert A., Adrian R. King, Ariana Lahijani, and Saad B Omer. 2019. “Current Landscape of Non-medical Vaccination Exemptions in the United States: Impact of Policy Changes.” *Expert Review of Vaccines* 18 (2): 175–90. <https://doi.org/10.1080/14760584.2019.1562344>.
- Bogart, N. 2021. “Do Vaccine Mandates Violate Canadians’ Charter Rights?” *CTV News*, September 1. <https://www.ctvnews.ca/lifestyle/article/do-vaccine-mandates-violate-canadians-charter-rights/>.
- Boyneburgk, Konrad v, and Francesca Bellazzi. 2022. “COVID-19 Vaccines and the Virtues.” *Public Health Ethics* 15 (3): 209–19. <https://doi.org/10.1093/phe/phac027>.
- Cairney, Paul. 2016. *The Politics of Evidence-Based Policy Making*. Palgrave Macmillan UK.
- Cave, Emma. 2017. “Voluntary Vaccination: The Pandemic Effect.” *Legal Studies* 37 (2): 279–304. <https://doi.org/10.1111/lest.12144>.
- Crawford, Blair. 2022. “Truck Convoy: Judge Adjourns Class Action Suit to Halt ‘Unbearable’ Air Horns.” *Ottawa Citizen*, February 5. <https://ottawacitizen.com/news/local-news/class-action-suit-claims-air-horns-causing-unbearable-torment>.
- Davidson, Peter. 2021. “Tony Blair Blasts Unvaccinated ‘Idiots’ as Fears Grow Over Spread of Omicron Variant.” *Daily Record*, December 22. <https://www.dailyrecord.co.uk/news/politics/tony-blair-blasts-unvaccinated-idiot-25762556>.
- EKOS Research Associates. 2018. “Survey for the Development of the Childhood Vaccination Campaign. Findings Report. Prepared for Health Canada.” <https://epe.lac-bac.gc.ca/100/200/301/pwgsc-tpsdc/por-ef/health/2018/022-17-e/report.pdf>.
- Fiore, B. 2021. “Mark McGowan Calls Anti-Vaxxers ‘Crazy and ‘Deranged’ in State Parliament.” *The West Australian*, November 16. <https://thewest.com.au/politics/state-politics/mark-mcgowan-calls-anti-vaxxers-crazy-and-deranged-in-state-parliament-c-4580801>.

- Friedman, Uri. 2017. "What Is a Populist?" *The Atlantic*, February 27. <https://www.theatlantic.com/international/archive/2017/02/what-is-populist-trump/516525>.
- Giubilini, Alberto. 2020. "An Argument for Compulsory Vaccination: The Taxation Analogy." *Journal of Applied Philosophy* 37 (3): 446–66. <https://doi.org/10.1111/japp.12400>.
- Global News. 2022. "COVID-19 Vaccine Mandates Take Centre Stage in Explosive House of Commons Exchange." *Global News*, February 7. <https://globalnews.ca/video/8601134/COVID-19vaccine-mandates-take-centre-stage-in-explosive-house-of-commons-exchange/>.
- Goldenberg, Maya J. 2016. "Public Misunderstanding of Science? Reframing the Problem of Vaccine Hesitancy." *Perspectives on Science* 24 (5): 552–81. https://doi.org/10.1162/POSC_a_00223.
- Goldenberg, Maya J. 2021. *Vaccine Hesitancy: Public Trust, Expertise, and the War on Science*. University of Pittsburgh Press.
- Goldenberg, Maya J. 2022. "Normative Theory and the COVID Pandemic: Author's Response to Miriam Solomon and Inmaculada de Melo-Martin." *IJFAB: International Journal of Feminist Approaches to Bioethics* 15 (2): 116–30. <https://doi.org/10.3138/ijfab.15.2.09>.
- Goldenberg, Maya J., Bipin Adhikari, Lorenz von Seidlein, Phaik Yeong Cheah, and Heidi J. Larson. 2023. "Vaccine Mandates and Public Trust Do Not Have to Be Antagonistic." *Nature Human Behaviour* 7 (10): 1605–6. <https://doi.org/10.1038/s41562-023-01720-8>.
- Government of Canada. 2020. "Regulating Vaccines for Human Use." September 9. <https://www.canada.ca/en/health-canada/services/drugs-health-products/biologics-radiopharmaceuticals-genetic-therapies/activities/fact-sheets/regulation-vaccines-human-canada.html>.
- Gravagna, Katie, Andy Becker, Robert Valeris-Chacin, et al. 2020. "Global Assessment of National Mandatory Vaccination Policies and Consequences of Non-Compliance." *Vaccine* 38 (49): 7865. <https://doi.org/10.1016/j.vaccine.2020.09.063>.
- Grossman, Lewis A. 2021. *Choose Your Medicine: Freedom of Therapeutic Choice in America*. Oxford University Press.
- Haelle, Tara. 2015. "Why the Not-So-Great Vaccine Debate of February 2015 Doesn't Matter." *Politico*, February 8. <https://www.politico.com/magazine/story/2015/02/vaccines-politics-115008/>.
- Hammersley, Martyn. 2005. "Is the Evidence-Based Practice Movement Doing More Good than Harm? Reflections on Iain Chalmers' Case for Research-Based Policy Making and Practice." *Evidence & Policy* 1 (1): 85–100. <https://doi.org/10.1332/1744264052703203>.
- Hodge, James G., Jr., and Lawrence O Gostin. 2002. "School Vaccination Requirements: Historical, Social, and Legal Perspectives." *Kentucky Law Journal* 90 (4): Article 7.
- Hoffman, Jan. 2020. "'I Won't Be Used as a Guinea Pig for White People'." *New York Times*, October 7. <https://www.nytimes.com/2020/10/07/health/coronavirus-vaccine-trials-african-americans.html>.
- Hudson, Deborah. 2022. "COVID-19 Vaccinations and Workplace Rights: 2022 Case Law Update." *Industrial Relations Centre – Queens University*. <https://irc.queensu.ca/COVID-19-vaccinations-and-workplace-rights-2022-case-law-update/>.
- Hwang, Tim. 2020. "Deconstructing the Disinformation War." *MediaWell, Social Science Research Council*, June 1. <https://mediawell.ssrc.org/expert-reflections/deconstructing-the-disinformation-war/>.
- Jamison, Amelia M., David A. Broniatowski, Mark Dredze, Zach Wood-Doughty, DureAden Khan, and Sandra Crouse Quinn. 2020. "Vaccine-Related Advertising in the Facebook Ad Archive." *Vaccine* 38 (3): 512–20. <https://doi.org/10.1016/j.vaccine.2019.10.066>.
- Krastev, Sekoul, Oren Krajden, Zoua M. Vang, et al. 2023. "Institutional Trust Is a Distinct Construct Related to Vaccine Hesitancy and Refusal." *BMC Public Health* 23 (1): 2481. <https://doi.org/10.1186/s12889-023-17345-5>.
- Marston, Greg, and Rob Watts. 2003. "Tampering with the Evidence: A Critical Appraisal of Evidence Based Policy." *The Drawing Board: An Australian Review of Public Affairs* 3 (3): 143–63.
- Matthews, R., and K. Menzel. 2023. "Vaccine Mandates and Cultural Safety." *Journal of Bioethical Inquiry* 20 (4): 719–30. <https://doi.org/10.1007/s11673-023-10319-7>.
- Mosby, Ian, and Jaris Swidrovich. 2021. "Medical Experimentation and the Roots of COVID-19 Vaccine Hesitancy among Indigenous Peoples in Canada." *CMAJ* 193 (11): E381–83. <https://doi.org/10.1503/cmaj.210112>.
- Naughtie, Andrew. 2022. "Trudeau Accuses Truckers of 'Hate, Abuse and Racism' as He Tests Positive for COVID." *The Independent*, January 31. <https://www.independent.co.uk/news/world/americas/justin-trudeau-covid-canada-convoy-b2004245.html>.
- Navin, Mark. 2013. "Resisting Moral Permissiveness about Vaccine Refusal." *Public Affairs Quarterly* 27 (1): 69–85. <https://www.jstor.org/stable/43574497>.

- Navin, Mark C., and Katie Attwell. 2019. "Vaccine Mandates, Value Pluralism, and Policy Diversity." *Bioethics* 33 (9): 1042–49. <https://doi.org/10.1111/bioe.12645>.
- Navin, Mark C., and Mark L. Largent. 2017. "Prioritizing Parental Liberty in Non-Medical Vaccine Exemption Policies: A Response to Giubilini, Douglas and Savulescu." *Public Health Ethics* 10 (3): 241–43. <https://doi.org/10.1093/phe/phx015>.
- Nogrady, Bianca. 2021. "I Hope You Die": How the COVID Pandemic Unleashed Attacks on Scientists." *Nature* 598 (7880): 250–53. <https://doi.org/10.1038/d41586-021-02741-x>.
- Office of the Privacy Commissioner of Canada. 2023. "Investigation into COVID-19 Vaccination Attestation Requirements Established by Certain Separate Employers of the Federal Public Service." May 30. https://www.priv.gc.ca/en/opc-actions-and-decisions/investigations/investigations-into-federal-institutions/2022-23/pa_20230529_pc/.
- Olive, Jacqueline K., Peter J. Hotez, Ashish Damania, and Melissa S. Nolan. 2018. "The State of the Antivaccine Movement in the United States: A Focused Examination of Nonmedical Exemptions in States and Counties." *PLOS Medicine* 15 (6): e1002578. <https://doi.org/10.1371/journal.pmed.1002578>.
- Omer, Saad B., Cornelia Betsch, and Julie Leask. 2019. "Mandate Vaccination with Care." *Nature* 571: 469–72.
- Omer, Saad B., Jennifer L. Richards, Michelle Ward, and Robert A. Bednarczyk. 2012. "Vaccination Policies and Rates of Exemption from Immunization, 2005–2011." *New England Journal of Medicine* 367: 1170–71. <https://doi.org/10.1056/NEJMc1209037>.
- Pierik, Roland. 2018. "Mandatory Vaccination: An Unqualified Defence." *Journal of Applied Philosophy* 35 (2): 381–97. <https://doi.org/10.1111/japp.12215>.
- Poland, Gregory A., and Robert M. Jacobson. 2012. "The Re-Emergence of Measles in Developed Countries: Time to Develop the Next-Generation Measles Vaccines?" *Vaccine* 30 (2): 103–4. <https://doi.org/10.1016/j.vaccine.2011.11.085>.
- Rainford, John, and Joshua Greenberg. 2015. "Taking Off the Gloves." *Policy Options*, July 6. <https://policyoptions.irpp.org/magazines/clearing-the-air/rainford-greenberg/>.
- Reiss, Dorit E., and Arthur L. Caplan. 2020. "Considerations in Mandating a New COVID-19 Vaccine in the USA for Children and Adults." *Journal of Law and the Biosciences* 7 (1): 1–9. <https://doi.org/10.1093/jlb/lsaa025>.
- Sarewitz, Daniel. 2004. "How Science Makes Environmental Controversies Worse." *Environmental Science & Policy* 7 (4): 385–403. <https://doi.org/10.1016/j.envsci.2004.06.001>.
- Schwartz, Jason L. 2012. "New Media, Old Messages: Themes in the History of Vaccine Hesitancy and Refusal." *AMA Journal of Ethics* 14 (1): 50–55. <https://doi.org/10.1001/virtualmentor.2012.14.1.mhst1-1201>.
- Ward, Jeremy K., James Colgrove, and Pierre Verger. 2018. "Why France Is Making Eight New Vaccines Mandatory." *Vaccine* 36 (14): 1801–3. <https://doi.org/10.1016/j.vaccine.2018.02.095>.
- WHO. 2022. "Coronavirus Disease (COVID-19) Advice for the Public: Mythbusters." *World Health Organization*, January 19. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public/myth-busters>.
- Williams, Daniel. 2023. "Misinformation Is the Symptom, Not the Disease." *IAI TV*, December 7. <https://iai.tv/articles/misinformation-is-the-symptom-not-the-disease-daniel-walliams-auid-2690>.
- Xinhua. 2021. "Romanian PM Likens Anti-Vaccine Campaigners to Terrorists." *Xinhuanet.com*, March 13. http://www.xinhuanet.com/english/2021-03/13/c_139806449.htm.

VALUES AND DUAL USE BIOMEDICAL RESEARCH

David B. Resnik

33.1 What is dual use research?

In a general sense, almost all scientific research is dual use because virtually all knowledge can be used for good or ill. The National Institutes of Health (NIH)'s (2025) dual use policy states that biological research is dual use if “the methodologies, materials, or results could be used to cause harm.” While this definition serves to raise awareness among scientists and the public about the dual nature of scientific research, it is too broad to offer any practical guidance because there are indefinitely many ways that methodologies, materials, or results could be used to cause harm ([Resnik 2021](#)). Accordingly, dual use research (DUR) policies adopted by US government agencies have used a narrower term, “dual use research of concern” (DURC), which the Biden Administration defined as:

[L]ife sciences research that, based on current understanding, can be reasonably anticipated to provide knowledge, information, products, or technologies that could be misapplied to do harm with no, or only minor, modification to pose a significant threat with potential consequences to public health and safety, agricultural crops and other plants, animals, the environment, materiel, or national security.

(The Whitehouse 2024)

This definition is similar to a definition adopted by the Department of Health and Human Services (DHHS) in 2021 ([Department of Health and Human Services 2021](#)). Other organizations with DUR policies use definitions similar to the US government definition or refer to the definition ([American Society of Microbiology 2023](#); [World Health Organization 2020](#)).

33.2 Historical and regulatory background

Although DUR has been conducted in some form for hundreds of years, a good place to begin our review of the history of DUR is in 1972, when 118 nations, including the US and Soviet Union, signed the Biological and Toxin Weapons Convention, which prohibits the production, stockpiling, acquisition, and deployment of biological weapons ([United Nations 2020](#)). Under the terms of the treaty, countries can only conduct defensive bioweapons research; however, the Soviet scientists continued

to conduct secret bioweapons research on the plague, anthrax, and Marburg and Ebola viruses until the breakup of the Soviet Union in the 1990s ([Leitenberg and Zilinskas 1989](#)). In 1979, 66 people died from anthrax poisoning as the result of an accident at a laboratory in Sverdlovsk, Russia ([Zhou, Song, and Wang 2019](#)).

The first known bioterrorist attack on US soil took place on September 17, 1984, when 750 people who ate at Taco Time in Dalles, OR, developed food poisoning. An investigation by the Centers for Disease Control and Prevention concluded that followers of the guru Bhagwan Shree Rajneesh had sprayed *Salmonella* bacteria in the salad bar in an attempt to sicken Dalles voters and influence local elections. Members of the religious cult had been planning to contaminate the town's water supply ([Resnik 2021](#)).

Concerns about bioterrorism became a major US government policy matter in October 2001, when Bruce Ivins, a microbiologist at the US Army Medical Research Institute of Infectious Diseases in Fort Detrick, allegedly sent letters laced with anthrax spores to Senators Tom Daschle and Patrick Leahy and several members of the media. Five people died as a result of the attacks, and dozens got sick. In 2008, the Federal Bureau of Investigation named Ivins as its chief suspect, but he committed suicide before he could be apprehended and brought to trial ([CNN 2023](#)).

The anthrax-laced letters were sent about a month after September 11, 2001, when nearly 3000 people died from terrorist attacks carried out by operatives working for Al-Qaeda. Although these attacks used commercial airplanes as weapons, US intelligence officials soon learned that Al-Qaeda had been trying to develop and deploy bioweapons since 1998 ([Juling 2023](#)). Although it is not known whether the person who mailed the anthrax-laced letters consulted scientific papers about anthrax, it became clear to scientists, intelligence officials, and policymakers that bioterrorism was an emerging threat.

In 2001–2002, the US Congress passed several laws designed to prevent future terrorist attacks by enhancing security, surveillance, and intelligence. Some of these laws established requirements for the possession, control, use, and transfer of dangerous biological entities known as select agents and toxins ([Young 2023](#)). Currently, there are 68 types of biological entities classified as select agents or toxins, including severe acute respiratory syndrome (SARS), SARS-associated coronaviruses, genetically manipulated forms of SARS, anthrax, Ebola, smallpox, plague, botulinum neurotoxin, ricin, and reconstructed pandemic influenza 1918 ([Federal Select Agent Program 2025](#)).

In 2002, several NIH-funded studies were published in top scientific journals that raised significant dual use concerns, including experiments showing how to increase virulence of the mousepox virus ([Jackson, Ramsay, and Christensen 2001](#)), how to increase virulence of a virus used to vaccinate people against smallpox ([Rosengard, Liu, and Nie 2002](#)), and how to manufacture a polio virus from DNA sequence data and mail-order supplies ([Cello et al. 2002](#)).

In response to these studies, Congress held hearings on using the results of biological research for bioterrorism and asked the National Research Council (NRC) to study the issue. In 2004, the NRC issued a report that examined the scientific, ethical, and policy issues raised by dangerous biological research. In the report, the NRC defined dual use, identified seven types of experiments of concern (mentioned above), and recommended that the US government form an agency to provide scientists, journal editors, and government officials with guidance and advice on biosecurity and biosafety issues ([National Research Council 2004](#)). In the fall of 2004, Congress acted on the NRC's recommendations and formed the National Science Advisory Board for Biosecurity (NSABB). The NSABB is an advisory body that includes members from microbiology, epidemiology, medicine, veterinary medicine, national security, agriculture, public health, ethics, and law ([Resnik 2013](#)).

Also in 2004, the DHHS learned that a paper had been submitted to the *Proceedings of the National Academies of Sciences* (PNAS) that described a mathematical model of the dispersal of botulinum toxin through the US milk supply. The paper predicted the number of people would die or become

seriously ill from different amounts of toxin introduced at various places in the supply chain (Wein and Liu 2005). DHHS asked the PNAS editors to not publish the paper, but the editors rejected this advice on the grounds that the benefits of publication would outweigh the risks. Although the editors recognized that information contained in the paper could be used to support an attack on the milk supply, they argued that the paper would help to alert public health officials and milk suppliers about vulnerabilities in the supply system so they would take proper steps to protect it (Alberts 2005; Resnik 2009).

During the mid-2000s, some journals began developing dual use review policies (Patrone, Resnik, and Chin 2012). For example, journals published by Nature Publishing Group require authors, reviewers, and editors to address any dual use concerns during the submission process and require special review for papers flagged as raising dual use issues. During the dual use review process, editors weigh the benefits and risks of publication. They may refuse to publish the paper, publish it only in redacted form (with risky information removed), or publish it in full, depending on how the risks and benefits are assessed (Nature 2025). Although some journals have dual use review policies, most do not (Resnik, Barner, and Dinse 2011).

In 2005, *Science* and *Nature* published two papers that described the genome of a reconstructed 1918 pandemic flu virus (Taubenberger, Reid, and Lourens 2005; Tumpey, Basler, and Aguilar 2005). While the papers were under review, DHHS asked the NSABB to render an advisory opinion concerning publication, and the NSABB unanimously recommended in favor of publishing (Kennedy 2005). This was the first test case for the newly formed board.

33.3 The controversy over H5N1 avian flu research

In 2011–2012, a major controversy concerning DUR erupted when two research groups, one led by Ronald Fouchier at Erasmus Medical Center in the Netherlands and the other by Yoshihiro Kawaoka at the University of Wisconsin-Madison, sought to publish the results from NIH-funded experiments demonstrating how to genetically modify the H5N1 avian influenza virus so that it can be transmitted between mammals through respiratory water droplets. The virus, which has a case fatality rate of 50%, had killed several hundred people but fortunately could only be acquired through direct contact with infected birds (Resnik 2013). The researchers conducted these dangerous experiments because they were concerned that H5N1 might naturally acquire mutations which allow it to be transmissible between mammals. Since influenza viruses and other pathogens periodically jump between species, this concern was a realistic possibility. Indeed, many infectious diseases in the human population, some of them deadly, have a zoonotic origin (Centers for Disease Control and Prevention 2021). Results from these studies, the researchers argued, could provide scientists with important knowledge about H5N1's structure and function that could be useful in the development of vaccines, treatments, and other countermeasures to protect the public from the threat posed by the virus's potential future evolution (Resnik 2013).

The authors submitted their papers to *Science* and *Nature* (Imai, Watanabe, and Hatta 2012; Russell, Fonville, and Brown 2012). The editors, who recognized that the results of the research could be readily used to make a bioweapon, asked the NSABB to review the research. At its December 2011 meeting, the NSABB recommended the papers should be published only if key information needed to replicate the experiments was redacted (Resnik 2013). The NSABB argued that the benefits of full publication did not outweigh the risks. The risks included biosecurity risks, that is, the risk that the research could be used for terrorist or other harmful purposes, and biosafety risks, that is, the risk that someone who tries to replicate the research might accidentally release a dangerous microbe, which could infect lab workers and the larger human population (Resnik 2021). The researchers submitted longer versions of their papers to the NSABB which included additional information about potential benefits of their research and biosecurity and biosafety risk management. At its March 2012 meeting,

the NSABB reversed its earlier ruling and recommended full publication of both papers. However, about one third of NSABB members felt that Fouchier's research should be published only in redacted form (Resnik 2013).

The DURC controversy continued to simmer in scientific journals, conferences, and NSABB meetings following the publication of these papers. The scholarly and scientific debate shifted from DURC in general to gain-of-function (GOF) research in virology, specifically. A GOF experiment in virology is an experiment designed to enhance the functioning of a virus, by for example, increasing its virulence, transmissibility, or host range (Selgelid 2016). The controversial H5N1 experiments were GOF experiments. While some scientists argued that the benefits of GOF outweigh the risks (Casadevall et al. 2014; Davis, Chen, and Pappas 2014; Fauci 2012; Fouchier 2015; Fouchier, García-Sastre, and Kawaoka 2012), others argued the opposite (Evans, Lipsitch, and Levinson 2015; Klotz 2015; Lipsitch 2018; Lipsitch and Inglesby 2014; Wain-Hobson 2013). However, both sides agreed that improvements in biosafety, biosecurity, and research oversight for GOF research were needed (Lipsitch and Bloom 2012; Casadevall et al. 2014).

33.4 DHHS policy on research involving potential pandemic pathogens

In October 2014, DHHS announced a temporary moratorium on funding of new research involving GOF experiments on potential pandemic pathogens (PPPs), such as SARS, Middle Eastern respiratory syndrome (MERS), and influenza. The purpose of the funding pause was to give the agency time to thoroughly review the scientific and ethical issues surrounding GOF experiments and to develop policies (Resnik 2021). DHHS ended the funding pause and announced a new policy framework for overseeing GOF research involving PPPs in December 2017 (Resnik 2021). The policy defines enhanced PPPs and establishes criteria for approving research involving enhanced PPPs, including (1) the research design must be scientifically sound; (2) the pathogen must be a credible source of a future human pandemic; (3) there are no feasible or efficacious alternative methods of addressing the research question; (4) the potential benefits of the research to society outweigh the potential risks; (5) the results will be responsibly communicated; (6) there will be appropriate institutional oversight of the research, including safety oversight; and (7) the research is justifiable in terms of ethical values, including beneficence, justice, respect for persons, and scientific freedom (Department of Health and Human Services 2017).

In 2014, researchers from the California Department of Public Health redacted some genomic sequence data from a paper on the genome of a novel strain of *Clostridium botulinum* (Dover, Barash, and Hill 2014) because they were concerned someone could use the data to make a toxin for which there was no effective anti-toxin (Relman 2014). They published the complete genome when the anti-toxin became available. This is believed to be one of the few well-known cases of redacted publication of DURC (Resnik 2021).

In 2015, an NIH-funded research team led by Ralph Baric at the University of North Carolina at Chapel Hill published a paper reporting the results of a GOF experiment involving a PPP. In these experiments, the researchers created a novel SARS coronavirus (SARS-CoV) that expresses the spike protein, SHC014, on its surface. The virus was able to use this spike protein to bind with angiotensin-converting enzyme 2 (ACE2) receptors to infect and replicate in human airway cells and caused pathogenesis in humanized mice. Because the genomic sequence for the spike protein was found in coronaviruses that infect horseshoe bat populations, the authors hypothesized that it is possible that a new form of SARS could emerge from horseshoe bats and infect the human population (Menachery, Yount, and Debbink 2015). Shi Zhengli, who provided the spike protein sequence, was a microbiologist at the Wuhan Institute of Virology (WIV). Zhengli had been conducting similar GOF experiments at the WIV. While Baric's lab was classified as a biosafety level (BSL)-3, Shi's was only BSL-2 (Jacobsen 2021). (BSL-4 is the safest type of lab.)

33.5 The COVID-19 pandemic

Although Baric's research did not generate much attention when it was published, it became controversial in May 2021, when US Senator Rand Paul questioned National Institute of Allergy and Infectious Diseases Director Anthony Fauci about the study during Senate Health Committee hearings. Baric's study gained attention because Senator Fauci was investigating the possibility that SARS-CoV-2, the virus that caused the COVID-19 pandemic, escaped from a lab at the WIV, as a result of GOF research funded by the NIH through the EcoHealth Alliance. Although Fauci initially denied that the NIH had funded GOF at the WIV, the NIH later admitted it had but was unaware of the research due to EcoHealth Alliance's failure to clearly report the use of its funds to the agency (Calvert and Arbutnutt 2023).

When the pandemic began in early 2020, US President Donald Trump claimed that intelligence sources indicated that the virus probably had escaped from the WIV, but scientists quickly moved to dismiss this notion as a conspiracy theory (Wade 2021). In February 2020, Fauci, NIH Director Frances Collins, and EcoHealth Alliance Director Peter Daszak organized an effort to quickly publish an article to prove that SARS-CoV-2 had a natural origin in an animal population and was not the product of a laboratory experiment (Ridley and Chan 2021). The authors presented several pieces of evidence for this conclusion (Andersen et al. 2020):

- Pathogens found in animal populations, especially corona and influenza viruses, often infect the human population; "natural spillover" events are common.
- SARS-CoV-2 is 96% genetically identical to a coronavirus found in horseshoe bats in China.
- The earliest cases of COVID-19 were associated with the Wet Market in Wuhan.
- Several animals sold at the Wet Market could have served as intermediate hosts for the virus.
- There were no telltale signs of genetic manipulation in the virus' genome.
- Although the virus could have been produced in a laboratory by serial passaging (a process in which human beings pass the virus between animals to select for viruses with enhanced functions) this would have required progenitor virus, which was not described in the literature.

Although this article expressed a high degree of confidence that COVID-19 was not a human-made virus, unredacted emails obtained from Freedom of Information Act (FOIA) requests have revealed that several scientists who drafted this article had serious doubts about the natural origin hypothesis. The emails also indicate that the scientists involved in pushing the natural origin hypothesis were concerned that public acceptance of the lab leak hypotheses would have significant political repercussions for relationships with China and could damage the reputation of the NIH because of its support the GOF research (Tobias 2023).

For about a year, the effort to brand the lab leak hypothesis as a conspiracy theory worked. The mainstream media bought the story, and scientists who defended the lab escape hypothesis were unable to publish their work in top journals and were told to keep quiet or even ridiculed by their peers (Quay 2023; Wade 2021). However, evidence against the natural origin hypothesis and in favor of the lab escape hypothesis began to emerge, including (Chen, Kalyar, and Chughtai 2024)

- Scientists had been unable to locate an intermediate animal host for SARS-CoV-2, despite a concerted effort to find one (Wade 2021).
- The history of the virus was not consistent with a natural origin, since zoonotic pathogens usually enter the human population at several different times and places, but it appears that SARS-CoV-2 entered the human population only once, in Wuhan, China.
- The virus was unusually pre-adapted to infect human beings because the spike protein on SARS-CoV-2 includes a furin cleavage site, which greatly enhances the virus's ability to infect human cells. Moreover, although virologists knew that the furin cleavage site could greatly enhance

infectiousness, the gene for this site was not present in SARS viruses circulating in horseshoe bat populations (Peacock, Goldhill, and Zhou 2021; Ridley and Chan 2021; Young 2023).

- There was considerable evidence, from the intelligence community and scientists with inside knowledge, that researchers at the WIV had been genetically manipulating coronaviruses and that several workers at WIV had been infected with illnesses with the same symptoms of SARS-CoV-2 in November 2019 (Young 2023).
- In early 2024, investigative journalists discovered, by means of FOIA requests, that in 2018 US Defense Advanced Research Projects Agency (DARPA) had reviewed a proposal to fund a research project to insert the gene for the furin cleavage site into SARS virus. The proposal was a blueprint for making a SARS virus almost identical to SARS-CoV-2. DARPA did not fund the proposal because it determined its risks did not outweigh its benefits (Reyes 2024).

As of today, the majority of US intelligence agencies have concluded that SARS-CoV-2 is a human-made virus that escaped from the WIV (The Whitehouse 2025a) even though scientists remain divided on the issue of laboratory vs. natural origin. While top scientists and politicians have called for an independent inquiry into the origins of the pandemic (Chen, Kalyar, and Chughtai 2024; Harrison and Sachs 2022; Reyes 2024; Young 2023), the origins of the virus may never be known with a high degree of confidence without the full cooperation of the Chinese government, which has undermined attempts by the World Health Organization and the US government to investigate the origins of the pandemic and determine whether the virus escaped from the WIV (Gregorian, Dilanian, and Winter 2023).

Regardless of whether SARS-Cov-2 originated in a laboratory or in nature, it is clear that the lab leak hypothesis is not, and never was, a conspiracy theory. Laboratory accidents sometimes occur, even when scientists are taking appropriate precautions (Resnik 2024; Young 2023). Indeed, in 2017 scientists and intelligence experts had expressed serious concerns about safety and security lapses in China's microbiology labs (Cyranoski 2017). The COVID-19 pandemic has inflamed the GOF/DURC controversy by alerting politicians and the public about dangerous experiments that occur in labs around the world, often with US funding (Young 2023). In November 2023, the House of Representatives passed a ban on federal funding of GOF involving PPPs (Kaiser 2023). However, the Senate did not pass the bill.

33.6 Biden Administration revises policy on research involving potential pandemic pathogens

On May 6, 2024, in response to mounting pressure from scientists, politicians, and citizens who were concerned about the risks of GOF research and DURC, the Biden Administration announced that it had revised federal policies on DURC and research involving PPPs (Kaiser 2024; The Whitehouse 2024). The revised policy expands the types of research that require review and oversight and expands the definition of a PPP to include a pathogen that could have pandemic potential "as a result of" being modified. The new 84-page policy includes the following elements:

- 1 The policy revises the definition of DURC (mentioned above).
- 2 There are two categories of DURC oversight, Category 1 and Category 2.
- 3 Category 1 is DURC that includes nine types of experiments of concern, similar to those identified by the NRC, involving dangerous biological agents and toxins. Category 1 research can be reviewed by the funding agency.
- 4 Category 2 is DURC that involves PPPs or PPEPs (potential pandemic enhanced pathogens). A PPP is defined as a "pathogen that is likely capable of wide and uncontrollable spread in a human

population and would likely cause moderate to severe disease and/or mortality in humans” ([The Whitehouse 2024](#)). Category 2 research includes 7 different types of experiment involving 15 types of biological agents and toxins. Category 2 requires funding agency and departmental level review, e.g., NIH and DHHS.

- 5 The policy provides guidance for investigators, institutions, and funding agencies concerning implementation of the policy.

33.7 Trump administration takes steps to tighten the control of gain-of-function research

On May 5, 2025, President Trump signed an executive order to tightly control GOF research ([The WhiteHouse 2025b](#)). The order requires the following:

- 1 Ends federal funding of dangerous GOF conducted in countries of concern (e.g., China) and other countries where there are not adequate biosafety standards.
- 2 Suspends federal funding for existing dangerous GOF research projects.
- 3 Instructs federal agencies to revise the Biden Administration’s DURC and PPP policies to strengthen oversight, accountability, and public transparency.
- 4 Instructs federal agencies to develop policies to manage the risks of non-federally funded GOF research.
- 5 Defines “dangerous GOF” as “scientific research on an infectious agent or toxin with the potential to cause disease by enhancing its pathogenicity or increasing its transmissibility” ([The WhiteHouse 2025b](#)). Although this definition is quite broad and covers theoretically applicable research designed to develop a vaccine for seasonal influenza, the executive order singles out activities that involve
 - a enhancing the harmful consequences of the agent or toxin;
 - b disrupting beneficial immunological response or the effectiveness of an immunization against the agent or toxin;
 - c conferring to the agent or toxin resistance to clinically or agriculturally useful prophylactic or therapeutic interventions against that agent or toxin or facilitating their ability to evade detection methodologies;
 - d increasing the stability, transmissibility, or the ability to disseminate the agent or toxin;
 - e altering the host range or tropism of the agent or toxin;
 - f enhancing the susceptibility of a human host population to the agent or toxin; or
 - g generating or reconstituting an eradicated or extinct agent or toxin ([The WhiteHouse 2025b](#)).

As of the writing of this chapter, federal government agencies have not fully implemented this executive order. In 2025, it is also likely that the US Congress will consider legislation to control GOF research ([Paul 2025](#)).

33.8 Social values and dual use research

Other chapters in the volume have developed and examined philosophical arguments for incorporating social values into different types of decisions related to funding, design, implementation, interpretation, and application of scientific research. It is not my aim in this chapter to engage in the science/values debate (for review of some of the key issues, see [Douglas \(2016\)](#); [Elliott \(2022\)](#)). Rather, I would like to describe how social values have influenced DURC while setting aside questions about the appropriateness of this influence.

Social values have influenced several types of issues related to DURC:

- 1 **Whether to conduct DURC.** Before funding issues ever arise, scientists must decide whether they want to conduct DURC and institutions must decide whether they will support it. Community members may also express their opinions about whether risky research should be performed in their neighborhood (Eaves 2020; Miller and Segelid 2007; Resnik 2021).
- 2 **Whether to fund DURC.** As discussed above, funding decisions have come up numerous times since the publication of the H5N1 GOF papers in 2012 and continuing through the COVID-19 pandemic.
- 3 **How to conduct DURC safely.** Even when there has been agreement about conducting and funding DURC, questions will still arise about how to conduct it safely, for example, how to prevent dangerous biological agents from escaping the lab (biosafety rules) or being stolen or diverted (select agent rules). This is an ethical issue, because measures designed to enhance safety can interfere with research and significantly increase costs. Thus, there are difficult questions about how safe is safe enough? Many have argued that the coronavirus research at the WIV was not safe enough (Ridely and Chan 2021).
- 4 **Whether and how to publish DURC.** As discussed above, publication issues have come up several times, for example, with the milk supply contamination study, the H5N1 GOF experiments and the research on the *C. botulinum* genome.

Underlying these issues is a meta-issue concerning DURC:

- 5 **Estimating the risks of DURC.** Estimates of the risks of DURC are strongly influenced by values, because they are based on models which depend on assumptions and these assumptions are motivated by values. The two main types of risks include biosecurity risks (that is, the risks that the research could be used to make a bioweapon) and biosafety risks (that is, the risks genetically engineered microbes might escape from the lab and infect the human population) (Resnik 2021). As noted above, there have been only two, well-documented bioterrorist attacks in the US, and in neither of these attacks did the perpetrators use the results of GOF research. The agents used in these attacks—anthrax and *Salmonella*—were not genetically engineered. Since there are no data that directly address the biosecurity risks of GOF research with PPPs, the best that one can do is to develop statistical models that estimate the risks of bioterrorism (Boddie, Watson, and Ackerman 2015; National Research Council 2008). However, these models are based on assumptions, which may be influenced by values, such as attitudes toward risk and attitudes toward GOF (Elliott 2022; Resnik 2021, 2024). While there are some data related to biosafety risks, these have to do with the risks of being infected by a biological agent during laboratory work, not the risk of transmitting that infection to the larger community (Kimman, Smit, and Klein 2008; Wurtz, Papa, and Hukic 2016). As a result, risk estimates for community transmission depend on assumptions. For example, Lipsitch and Inglesby (2014), who are microbiologists and are opposed to GOF, estimated the risk that this research would infect the larger community to be between 0.05% and 0.6% per laboratory worker per year. However, Fouchier (2015), who is in favor of GOF research and has received NIH funding for GOF research, estimated the risk to be several orders of magnitude less than Lipsitch and Inglesby's estimate.

Some of the key values at stake in debates about DURC include the following:

- **Public health.** Proponents and opponents of DURC have both appealed to the value of public health. For example, proponents of GOF research involving PPPs have argued that these experiments are important for promoting public health because they can produce knowledge that could

be used to develop vaccine and treatments that protect humanity from emerging deadly diseases (Fauci 2012), but opponents have argued that these experiments threaten public health because they involve creating dangerous microbes that could escape from the lab and trigger a pandemic or could be used as bioweapons (Lipsitch 2018).

- **National security.** Some commentators have argued that DURC should be kept secret or published only in redacted form to protect national security (Resnik 2009). This was a key issue in the controversy surrounding Wein and Liu's (2005) paper on infecting the US milk supply and the H5N1 GOF experiments. National security concerns played a key role in Dover, Barash, and Hill's (2014) decision to redact some data from their paper on the genome of novel strain of *C. botulinum*. Others have argued, however, that keeping DURC secret threatens democracy (see below).
- **Progress of science, freedom of inquiry, and scientific openness.** Commentators have appealed to these values to argue against restrictions on DURC, because restrictions on it can interfere with values essential to the progress of science, such as freedom of inquiry and scientific openness and sharing of information (Resnik 2013; Selgelid 2016).
- **Democracy and transparency.** Commentators have appealed to these values to argue for informing the public about government-supported DURC and allowing the public to make decisions about DURC funding and oversight (Evans, Lipsitch, and Levinson 2015; Resnik 2021). Concerns about democracy and transparency have become more important as the public become increasingly concerned about risky biological research (Eaves 2020; Resnik 2021; Subbaraman 2020). Others have argued, however, that secrecy in some cases is necessary to protect national security and peer review process.
- **International politics.** Scientists who sought to brand the lab escape hypothesis concerning the origins of SARS-CoV-2 as a conspiracy theory were concerned about how public acceptance of the hypothesis would impact relations with China.
- **Domestic politics.** Many people viewed the lab leak hypothesis as a conspiracy theory because they were politically opposed to President Trump, who had suggested that SARS-CoV-2 may have escaped from the WIV. Indeed, the lab escape hypothesis became associated with support for President Trump and greater openness to examining the lab escape hypothesis did not occur until he left office (Wade 2021).
- **Public support for science.** Opponents of the lab escape hypothesis concerning the origins of SARS-CoV-2 were worried that public acceptance of this hypothesis would undermine public support for science and the NIH in particular, since it would imply that scientists responsible for starting a pandemic that has killed over 7 million people (Wade 2021).

33.9 Conclusion

Because DURC has significant implications for public health, national security, and democracy, it is an area of science where there is little debate about the importance of social values in overseeing the conduct, funding, and publication of research. It is likely that dilemmas and debates concerning DURC will continue to arise as scientists explore new areas of inquiry and develop new technologies. Hopefully, government officials, journal editors, university and business leaders, and other key decision-makers will be able to respond to these controversies in ways that will protect the interests of science and society. Because the issues are complex and involve conflicts between competing values, we should not expect to find simple and easy answers to the ethical and policy questions concerning DURC. However, in my opinion, processes that promote democratic deliberation are essential to making progress toward fair and effective solutions (Resnik 2024). Some of these include the following: (1) public, community, and stakeholder engagement concerning DURC, (2) community

representation on institutional biosafety committees, (3) public membership on committees that review government-funded DURC research proposals, and (4) enhanced transparency and openness concerning DURC research.

Acknowledgments

This research was supported by the intramural program of the NIH. It does not represent the views of the NIH, DHHS or US government.

References

- Alberts, Bruce. 2005. "Modeling Attacks on the Food Supply." *Proceedings of the National Academies of Sciences* 102 (28): 9737–8. <https://doi.org/10.1073/pnas.0504944102>.
- American Society of Microbiology. 2023. Dual Use Research. <https://journals.asm.org/dual-use-research>.
- Andersen, Kristian G., Andrew Rambaut, Walter Ian Lipkin, Edward C. Holmes, and Robert F. Garry. 2020. "The Proximal Origin of SARS-CoV-2." *Nature Medicine* 26: 450–2. <https://doi.org/10.1038/s41591-020-0820-9>.
- Boddie, Crystal, Matthew Watson, Gary Ackerman, et al. 2015. "Biosecurity. Assessing the Bioweapons Threat." *Science* 349 (6250): 792–3. <https://doi.org/10.1126/science.aab0713>.
- Calvert, Jonathan, and George Arbuthnutt. 2023. "What Really Happened Inside the Wuhan Lab Weeks Before Covid Erupted." *The Times*, June 10. <https://www.thetimes.co.uk/article/inside-wuhan-lab-covid-pandemic-china-america-qhjwwvm0>.
- Casadevall, Arturo, Don Howard, and Michael J. Imperiale. 2014. "An Epistemological Perspective on the Value of Gain-of-Function Experiments Involving Pathogens with Pandemic Potential." *mBio* 5 (5): e01875-14. <https://doi.org/10.1128/mBio.01875-14>.
- Cello, Jeronimo, Aniko V. Paul, and Eckard Wimmer. 2002. "Chemical Synthesis of Poliovirus cDNA: Generation of Infectious Virus in the Absence of Natural Template." *Science* 297 (5583): 1016–8. <https://doi.org/10.1126/science.1072266>.
- Centers for Disease Control and Prevention. 2021. "Zoonotic Diseases." <https://www.cdc.gov/onehealth/basics/zoonotic-diseases.html>.
- Chen, Xin, Fatema Kalyar, Abrar A. Chughtai, and Chandini McIntyre. 2024. "Use of a Risk Assessment Tool to Determine the Origin of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)." *Risk Analysis* 44: 1896–906. <https://doi.org/10.1111/risa.14291>.
- CNN. 2023. "2001 Anthrax Attacks: Fast Facts." <https://www.cnn.com/2013/08/23/health/anthrax-fast-facts/index.html>.
- Cyranoski, David. 2017. "Inside the Chinese Lab Poised to Study World's Most Dangerous Pathogens." *Nature* 542 (7642): 399–400. <https://doi.org/10.1038/nature.2017.21487>.
- Davis, C. Todd, Li-Mei Chen, Claudia Pappas, et al. 2014. "Use of Highly Pathogenic Avian Influenza A(H5N1) Gain-of-Function Studies for Molecular-Based Surveillance and Pandemic Preparedness." *mBio* 5 (6): e02431-14. <https://doi.org/10.1128/mBio.02431-14>.
- Department of Health and Human Services. 2017. "Framework for Guiding Funding Decisions about Proposed Research Involving Enhanced Potential Pandemic Pathogens." <https://www.phe.gov/s3/dualuse/Documents/P3CO.pdf>.
- Department of Health and Human Services. 2021. "Dual Use Research of Concern." <https://www.phe.gov/s3/dualuse/Pages/default.aspx>.
- Douglas, Heather. 2016. "Values in Science." In *The Oxford Handbook of Philosophy of Science*, edited by Paul Humphreys. Oxford University Press.
- Dover, Nir, Jason R. Barash, Karen K. Hill, Gary Xie, and Stephen S. Arnon. 2014. "Molecular Characterization of a Novel Botulinum Neurotoxin Type H Gene." *Journal of Infectious Diseases* 209 (2): 192–202. <https://doi.org/10.1093/infdis/jit450>.
- Eaves, Elisabeth. 2020. "The Risks of Building Too Many Bio Labs." *The New Yorker*, March 18, 2020. <https://www.newyorker.com/science/elements/the-risks-of-building-too-many-bio-labs>.
- Elliott, Kevin. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Evans, Nicholas G., Marc Lipsitch, and Meira Levinson. 2015. "The Ethics of Biosafety Considerations in Gain-of-Function Research Resulting in the Creation of Potential Pandemic Pathogens." *Journal of Medical Ethics* 41 (11): 901–8. <https://doi.org/10.1136/medethics-2014-102619>.

- Fauci, Anthony S. 2012. "Research on Highly Pathogenic H5N1 Influenza Virus: The Way Forward." *mBio* 3 (5): e00359-12. <https://doi.org/10.1128/mBio.00359-12>.
- Federal Select Agent Program. 2025. *Select Agents and Toxins*. January 14. <https://www.selectagents.gov/sat/list.htm>.
- Fouchier, Ron A. 2015. "Studies on Influenza Virus Transmission between Ferrets: The Public Health Risks Revisited." *mBio* 6 (1): e02560-14. <https://doi.org/10.1128/mBio.02560-14>.
- Fouchier, Ron A., Aldolfo García-Sastre, and Yoshiro Kawaoka. 2012. "The Pause on Avian H5N1 Influenza Virus Transmission Research Should Be Ended." *mBio* 3 (5): e00358-12. <https://doi.org/10.1128/mBio.00358-12>.
- Gregorian, Dareh, Ken Dilanian, Tom Winter, and Sahil Kapur. 2023. "FBI Director Accuses China of Trying to 'Thwart and Obfuscate' Covid Origin Probe." *NBC News*, February 28. <https://www.nbcnews.com/politics/politics-news/fbi-director-says-china-trying-thwart-covid-origin-probe-rcna72805>.
- Harrison, Neil L., and Jeffrey D. Sachs. 2022. "A Call for an Independent Inquiry into the Origin of the SARS-CoV-2 Virus." *Proceedings of the National Academies of Science* 119 (21): e2202769119. <https://doi.org/10.1073/pnas.2202769119>.
- Imai, Masaki, Tokiko Watanabe, Masato Hatta, et al. 2012. "Experimental Adaptation of an Influenza H5 HA Confers Respiratory Droplet Transmission to a Reassortant H5 HA/H1N1 Virus in Ferrets." *Nature* 486 (7403): 420–8. <https://doi.org/10.1038/nature10831>.
- Jackson, Ronald J., Alistair J. Ramsay, Christina D. Christensen, Sandra Beaton, Diana F. Hall, and Ian A. Ramshaw. 2001. "Expression of Mouse Interleukin-4 by a Recombinant Ectromelia Virus Suppresses Cytolytic Lymphocyte Responses and Overcomes Genetic Resistance to Mousepox." *Journal of Virology* 75 (3): 1205–10. <https://doi.org/10.1128/jvi.75.3.1205-1210.2001>.
- Jacobsen, Rowan. 2021. "Inside the Risky Bat-Virus Engineering that Links America to Wuhan." *MIT Technology Review*, June 29. <https://www.technologyreview.com/2021/06/29/1027290/gain-of-function-risky-bat-virus-engineering-links-america-to-wuhan/>.
- Juling, Dominik. 2023. "Future Bioterror and Biowarfare Threats for NATO's Armed Forces Until 2030." *Marine Corps University, Jams* 14 (1). <https://doi.org/10.21140/mcu.20231401005>.
- Kaiser, Jocelyn. 2023. "House Approves Ban on Gain-of-Function Pathogen Research." *Science*, November 15. <https://www.science.org/content/article/house-approves-ban-gain-function-pathogen-research>.
- Kaiser, Jocelyn. 2024. "White House Overhauls Rules for Risky Pathogen Studies." *Science*, May 7. https://www.science.org/content/article/white-house-overhauls-rules-risky-pathogen-studies?utm_source=sfmc&utm_medium=email&utm_content=alert&utm_campaign=SCleToc&et rid=35371335&et_cid=5207966.
- Kennedy, Donald. 2005. "Better Never Than Late." *Science* 310 (5746): 195. <https://doi.org/10.1126/science.310.5746.195>.
- Kimman, Tjeerd G., Eric Smit, and Michèl R Klein. 2008. "Evidence-Based Biosafety: A Review of the Principles and Effectiveness of Microbiological Containment Measures." *Clinical Microbiology Review* (3): 403–25. <https://doi.org/10.1128/CMR.00014-08>.
- Klotz, Lynn C. 2015. "Comments on Fouchier's Calculation of Risk and Elapsed Time for Escape of a Laboratory-Acquired Infection from His Laboratory." *mBio* 6 (2): e00268-15. <https://doi.org/10.1128/mBio.00268-15>.
- Leitenberg, Milton, and Raymond Zilinskas. 1989. *The Soviet Biological Weapons Program: A History*. Harvard University Press.
- Lipsitch, Marc. 2018. "Why Do Exceptionally Dangerous Gain-of-Function Experiments in Influenza." *Methods of Molecular Biology* 1836: 589–608. https://doi.org/10.1007/978-1-4939-8678-1_29.
- Lipsitch, Marc, and Barry R. Bloom. 2012. "Rethinking Biosafety in Research on Potential Pandemic Pathogens." *mBio* 3 (5): e00360-12. <https://doi.org/10.1128/mBio.00360-12>.
- Lipsitch, Marc, and Thomas V. Inglesby. 2014. "Moratorium on Research Intended to Create Novel Potential Pandemic Pathogens." *mBio* 5 (6): e02366-14. <https://doi.org/10.1128/mBio.02534-14>.
- Menachery, Vineet D., Boyd L. Yount Jr, Kari Debink, et al. 2015. "A SARS-Like Cluster of Circulating Bat Coronaviruses Shows Potential for Human Emergence." *Nature Medicine* 21: 1508–13. <https://doi.org/10.1038/nm.3985>.
- Miller, Seumas, and Michael J. Selgelid. 2007. "Ethical and Philosophical Consideration of the Dual-Use Dilemma in the Biological Sciences." *Science and Engineering Ethics* 13 (4): 523–80. <https://doi.org/10.1007/s11948-007-9043-4>.
- National Institutes of Health. 2025. Dual-use Research. February 14, 2025. <https://oir.nih.gov/sourcebook/ethical-conduct/special-research-considerations/dual-use-research>.
- National Research Council. 2004. *Biotechnology Research in an Age of Terrorism: Confronting the Dual Use Dilemma*. National Academies Press.

- National Research Council. 2008. *Department of Homeland Security Bioterrorism Risk Assessment*. National Academies Press.
- Nature. 2025. Research Ethics. <https://www.nature.com/nature-portfolio/editorial-policies/ethics-and-biosecurity>.
- Patrone, Daniel, David B. Resnik, and Lisa Chin. 2012. "Biosecurity and the Review and Publication of Dual-Use Research of Concern." *Biosecurity and Bioterrorism* 10 (3): 290–8. <https://doi.org/10.1089/bsp.2012.0011>.
- Paul, Rand. 2025. Risky Research Review Act. <https://www.paul.senate.gov/wp-content/uploads/2025/03/Risky-Research-Act-of-2025.pdf>.
- Peacock, Thomas P., Daniel H. Goldhill, Jie Zhou, et al. 2021. "The Furin Cleavage Site in the SARS-CoV-2 Spike Protein Is Required for Transmission in Ferrets." *Nature Microbiology* 6 (7): 899–909. <https://doi.org/10.1038/s41564-021-00908-w>.
- Quay, Stephen. 2023. "I Suspected a China Lab Leak Early On—But my Research Was Rejected." *Newsweek*. March 21. <https://www.newsweek.com/covid-lab-leak-china-virus-nuclear-war-1787390>.
- Relman, David A. 2014. "'Inconvenient Truths' in the Pursuit of Scientific Knowledge and Public Health." *Journal of Infectious Disease* 209 (2): 170–2. <https://doi.org/10.1093/infdis/jit529>.
- Resnik, David B. 2009. *Playing Politics with Science: Balancing Scientific Independence and Public Oversight*. Oxford University Press.
- Resnik, David B. 2013. "H5N1 Avian Flu Research and the Ethics of Knowledge." *Hastings Center Report* 43 (2): 22–33. <https://doi.org/10.1002/hast.143>.
- Resnik, David B. 2021. *Precautionary Reasoning in Environmental and Public Health Policy*. Springer Nature.
- Resnik, David B. 2024. "Biosafety, Biosecurity, and Bioethics." *Monash Bioethics Review*, July 30, 2024. <https://doi.org/10.1007/s40592-024-00204-3>.
- Resnik, David B., Dionne D. Barner, and Gregg E. Dinse. 2011. "Dual-Use Review Policies of Biomedical Research Journals." *Biosecurity and Bioterrorism* 9 (1): 49–54. <https://doi.org/10.1089/bsp.2010.0067>.
- Reyes, Ronny. 2024. "Scientist Claims 'Smoking Gun' Evidence COVID-19 Intentionally Created by Researchers in Chinese Lab." *New York Post*, February 29. <https://nypost.com/2024/02/29/world-news/scientists-may-have-started-the-covid-pandemic-article/>.
- Ridley, Matt, and Emily Chan. 2021. *Viral: The Search for the Origins of COVID-19*. Harper.
- Rosengard, Ariella, Yu Liu, Zhiping Nie, and Robert Jimenez. 2002. "Variola Virus Immune Evasion Design: Expression of a Highly Efficient Inhibitor of Human Complement." *Proceedings of the National Academies of Science* 99 (13): 8808–13. <https://doi.org/10.1073/pnas.112220499>.
- Russell, Colin A., Judith M. Fonville, André Brown, et al. 2012. "The Potential for Respiratory Droplet-Transmissible A/H5N1 Influenza Virus to Evolve in a Mammalian Host." *Science* 336 (6088): 1541–7. <https://doi.org/10.1126/science.1222526>.
- Selgelid, Michale J. 2016. "Gain-of-Function Research: Ethical Analysis." *Science and Engineering Ethics* 22 (4): 923–64. <https://doi.org/10.1007/s11948-016-9810-1>.
- Subbaraman, Nidhi. 2020. "US Officials Revisit Rules for Disclosing Risky Disease Experiments." *Nature*, January 27. <https://www.nature.com/articles/d41586-020-00210-5>.
- Taubenberger, Jeffery, Ann Reid, Raina Lourens, Ruixue Wang, Guozhong Jin, and Thomas G. Fanning. 2005. "Characterization of the 1918 Influenza Virus Polymerase Genes." *Nature* 437 (7060): 889–93. <https://doi.org/10.1038/nature04230>.
- The Whitehouse. 2024. Implementation Guidance for the United States Government Policy for Oversight of Dual Use Research of Concern and Pathogens with Enhanced Pandemic Potential. <https://www.whitehouse.gov/wp-content/uploads/2024/05/USG-DURC-PEPP-Implementation-Guidance.pdf>.
- The Whitehouse. 2025a. Lab Leak: The True Origins of COVID-19. <https://www.whitehouse.gov/lab-leak-true-origins-of-covid-19/>.
- The Whitehouse. 2025b. Improving the Safety and Security of Biological Research. <https://www.whitehouse.gov/presidential-actions/2025/05/improving-the-safety-and-security-of-biological-research/>.
- Tobias, James. 2023. "Evolution of a Theory. Unredacted Emails Shows Efforts to Rule out Lab Origin of COVID." *The Intercept*, January 19. <https://theintercept.com/2023/01/19/covid-origin-nih-emails/>.
- Tumpey, Terrence M., Christopher F. Basler, Patricia V. Aguilar, et al. 2005. "Characterization of the Reconstructed 1918 Spanish Influenza Pandemic Virus." *Science* 310 (5745): 77–80. <https://doi.org/10.1126/science.1119392>.
- United Nations. 2020. Convention on the Prohibition of the Development, Production and Stockpiling of Bacteriological (Biological) and Toxin Weapons and on their Destruction. <https://legal.un.org/avl/ha/cpdpsbttwd/cpdpsbttwd.html>.
- Wade, Nicholas. 2021. "The Origin of COVID: Did People or Nature Open Pandora's Box at Wuhan?" *Bulletin of Atomic Scientists*, May 5. <https://thebulletin.org/2021/05/the-origin-of-covid-did-people-or-nature-open-pandoras-box-at-wuhan/>.

- Wain-Hobson, Simon. 2013. "H5N1 Viral-Engineering Dangers Will Not Go Away." *Nature* 495 (7442): 411. <https://doi.org/10.1038/495411a>.
- Wein, Lawrence, and Yifan Liu. 2005. "Analyzing a Bioterror Attack on the Food Supply: The Case of Botulinum Toxin in Milk." *Proceedings of the National Academy of Sciences* 102 (28): 9984–9. <https://doi.org/10.1073/pnas.0408526102>.
- World Health Organization. 2020. What is Dual Use Research of Concern? <https://www.who.int/news-room/questions-and-answers/item/what-is-dual-use-research-of-concern>.
- Wurtz, N., A. Papa, M. Hukic, et al. 2016. "Survey of Laboratory-Acquired Infections Around the World in Biosafety Level 3 and 4 Laboratories." *European Journal of Clinical Microbiology and Infectious Disease* 35 (8): 1247–58. <https://doi.org/10.1007/s10096-016-2657-1>.
- Young, Allison. 2023. *Pandora's Gamble: Lab Leaks, Pandemics, and the World at Risk*. Center Street.
- Zhou, Dongsheng, Hongbin Song, Jianwei Wang, et al. 2019. "Biosafety and Biosecurity." *Journal of Biosafety and Biosecurity* 1 (1): 15–8. <https://doi.org/10.1016/j.jobbb.2019.01.001>.

PART V

Values in Scientific Activities



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

34

VALUES AND MEASUREMENT

Cristian Larroulet Philippi

34.1 Introduction

If anything in science is value-free, one might think, it is measurement. When we measure an object—unlike when we model a phenomenon or contemplate the possible applications of some scientific result—we are merely letting nature speak. Max Planck famously said, “an experiment is a question which science poses to nature, and a measurement is the recording of nature’s answer” (1949, 325).

If measurement is merely recording nature’s answer, there is little space for an analysis in which social, ethical, or political factors and considerations (“values” henceforth) play a significant role. “But before an experiment can be performed,” continued Planck, “it must be *planned*—the question to nature must be *formulated* before being posed. Before the result of a measurement can be used, it must be *interpreted*—nature’s answer must be understood properly” (Planck 1949, emphases added). That is, “recording nature’s answer” provides an impoverished understanding of what measurement involves. Once we adopt richer understandings, do values play significant roles? Which roles? And why, exactly?

This chapter introduces prominent arguments for the claim that values play a role in measurement practices. As befits a short handbook chapter, I aim for a wider coverage (versus in-depth analyses) of arguments. Accordingly, “values” and “play a role” will mean different things across various arguments. (For useful conceptual distinctions, see Elliott and Korf (2024) and Ward (2021).)

Some arguments concern basic background assumptions behind measurement (Section 34.2), others emphasize the consequences of measurement practices (Section 34.3), while still others consider the content of the concepts under measurement (Section 34.4). Admittedly, this distinction between consequences and content is not sharp (e.g., some normatively salient *consequences* are consequences of decisions concerning the *content* of measurements), but it remains useful for organizing the discussion. By emphasizing links between the values-in-science and the measurement literatures, I aim to show the need for closer engagement and highlight important open questions.

34.2 Background assumptions behind measurement

Consider some of the broad, background assumptions that motivate and shape the character of measurement practices.

“If something exists, it exists in some definite amount, and it can therefore be measured precisely” is an old (Pythagorean) idea (Michell 1999). But it has *not* shaped scientific practices always,

everywhere. Indeed, historians study “when did nature acquire sharp edges?”, that is, “when, why, and how *it began to make sense* to scientists, bureaucrats, merchants, and manufacturers to strive for ever more decimal places, ever more exact and exacting measurements, in the conviction that the mills of nature ‘grind exceeding small’” (Daston 1996, 517, emphasis added). Sharp edges, which make increases in precision in principle always possible; internally uniform categories, which make classifying and counting in principle unproblematic; homogeneous causal forces that can be measured quantitatively thanks to simple universal laws—it has *not* always made sense to conceive of natural (let alone social) phenomena in these ways. For example, while in the seventeenth century Robert Hooke didn’t think of natural phenomena as exact enough to motivate the search for higher and higher precision, a century later Lavoisier was reporting weights with eight decimal places (Daston 1996; see also Koyré (1953) and Wise (1995)). For a more radical contrast, consider the romantic movement: It decried the idea, endorsed by Kepler among others, that a geometer created our world—it was a poet!—and accordingly rejected quantity while embracing quality (Berlin 2013, 56).

After the successes in astronomy and geodesy in previous centuries, the late nineteenth century saw scientists in other fields striving for precise quantitative measurement. These efforts gave rise to “metrology”—the science of measurement, in charge of establishing standard units, which required setting up international networks across the globe. Historians argue that we cannot understand these efforts and successes independently of large-scale cultural and political transformations like industrialization, factory-like mentalities in science, processes of empire building, and large-scale military projects (e.g., O’Connell 1993; Porter 1995; Schaffer 1988; Wise 1995). For instance, Wise and Smith (1986) claim that industrialist values shaped William Thomson’s (Lord Kelvin) theoretical and metrological work. Such values, they argue, influenced both the central function he attributed to measurement and the “character of the quantities measured” (148).

That broad (cultural, metaphysical, etc.) assumptions shape measurement practices is more transparent in the human sciences. Consider the controversial judgment that conceptualizing “intelligence” as a measurable quantity “makes sense,” i.e., that it doesn’t do (too much) violence to the underlying phenomena. People, the idea goes, can reasonably be thought as having relatively stable amounts of intelligence (e.g., Spearman’s “mental energy”), amounts that we measure using intelligence tests just as we measure amounts of temperature using mercury-filled tubes. Something like this idea has shaped psychometric practice (Michell 1999; Wijsen, Borsboom, and Alexandrova 2022).¹ But it is rejected by the many who think of human attributes like intelligence as “processes that emerge from a situation and are part of the relationships of that situation” and not as relatively fixed properties of individuals (Rose, Lewontin, and Kamin 1984, 148). This deep disagreement is not easily explained without bringing in different (epistemological, metaphysical) broad background assumptions.

Similarly, while various authors agree that a *unidimensional* concept of intelligence expresses a *moral* commitment, their take on such a commitment differs dramatically: Wijsen, Borsboom, and Alexandrova (2022) link one-dimensionality to *equality*, because the former renders possible only differences in *degrees* of intelligence—not in *kinds* or absolute differences. We all have the same kind of intelligence, but some have more of it; hence drawing categorical divisions between groups in terms of intelligence is arbitrary, the idea goes. In contrast, other authors think that conceiving intelligence one-dimensionally reflects a *hierarchical* view of society, since it makes it impossible to represent different people as being more cognitively able in different dimensions, avoiding a ranking (Anderson 2002).

34.3 Consequences of measurement

As in other discussions of values-in-science (e.g., see Ward (2021) on values as “affected goods”), scholars highlight the *socially significant consequences* that measurement practices can have. Especially when these consequences are traceable to (epistemically) *unforced* choices (“contingent” in

Brown (2020, chap. 2)), many think scientists should bring in non-epistemic criteria to justify such choices. I'll sketch four arguments focusing on consequences; then I'll introduce two noteworthy literatures (one recent, one by practitioners) grappling with the consequences of measurements.

34.3.1 *What's being displaced*

One consequence worth considering is the kind of research that is displaced by the pursuit of quantification. Historian Deborah Coen (2013) offers an interesting example. During the nineteenth century, some seismological research on the nature and effects of earthquakes involved the active participation of the public. Swiss citizens, for example, were crucial agents in the production of seismological reports, which were rich in local details, sensitive to both scientific and local residents' concerns, led to important discoveries (e.g., the location of geological faults) and provided the background for the development of early scales for measuring earthquakes (e.g., the Mercalli scale).

The turn of the century saw a shift toward a "global seismology," with a much stronger reliance on scientific instruments (seismometers) than on people's reports, and a relative neglect of local patterns of variations in favor of large-scale ones. This culminated in the development of so-called magnitude scales (e.g., Richter's scale published in 1935) that don't depend on people's reports and aimed to measure the physical cause (a release of energy) of an earthquake rather than its various physical and human consequences at Earth's surface. This more global and quantitative research program no doubt delivered many scientific and policy-making achievements. But Coen's study invites us to acknowledge that it also displaced valuable scientific practices, practices that "aimed to enlist the public in the *coproduction* of knowledge" and at promoting "popular enlightenment" (2013, 79, author's emphasis).

34.3.2 *The consequences of metrological standards*

A crucial methodological choice in quantitative measurement is the selection of metrological standards, which fix when two magnitudes are considered equal. For example, Chang (2004, chap. 2) narrates early thermometry's debates about which fluid to use as congruence criterion for measuring temperature: Alcohol, mercury, or air? This is commonly discussed as a purely epistemic issue, i.e., as the search for *the* epistemically sanctioned standard.

How epistemically forced the choice of standards is, however, remains disputed within philosophy of measurement. Famously, conventionalists (e.g., Poincaré, Carnap) thought that it is not completely forced. There may be various epistemically sanctioned standards, and among them physicists choose the one making the laws of physics simpler. But if the choice isn't forced, why include only considerations of simplicity? Why not consider potential social consequences as well? For instance, given that many scientific and practical activities well beyond physical research require thermometers, there may be relevant non-epistemic consequences of this choice. Say, choosing mercury (over alcohol/air) economically benefits whoever has access to that metal. This is a straightforward consequence that could raise normative concerns, depending on various circumstances.

Still, since methodological choices will likely have social consequences at some point down the road, this might be thought of as a "trivial" or uninteresting point (cf. Ward 2021, 57). Given the enormous role that measurement standards play in science and trade globally—and hence the political-economic interests at stake in metrological choices—it's hard to consider the issue as trivial. Detailed studies of metrological practices indeed reveal that political considerations end up influencing things as technical as the calculation of coordinated universal time (Tal 2016) or the selection of electric standards (O'Connell 1993)—precisely because so much is at stake. Just think that large trade wars have been triggered by differences in measurement procedures (Källgren et al. 2003).

An interesting kind of case is when units (e.g., the meter, the kilogram) are no longer defined with concrete artifacts (a prototype bar, a cylinder of platinum-iridium kept near Paris) but “intrinsically” (using the speed of light, the Planck constant). These transitions, common during the last decades, can involve significant redistributions (decentralization) of authority and power across countries and institutions, with resulting political and economic consequences (O’Connell 1993). So far, philosophers have largely analyzed the selection of metrological standards only in *epistemic* terms, leaving to historians the value-based analyses; Tal’s (2016) is a notable exception.

34.3.3 *The consequences of error*

Because many measurements feed directly into decision-making, the treatment of *uncertainty* is a central theme in the metrological literature. Consider the practice called “conformity assessment” or “inspection” (e.g., Joint Committee for Guides in Metrology (JCGM) 2012; Källgren et al. 2003; Lira 2002).² Here, measurement results are used to decide whether an object (e.g., a sample of industrial waste water) conforms to a “tolerance interval” (e.g., a range within which the level of concentration of some pollutant in a sample of water must lie). When this inspection produces a binary decision—accept/reject the item—we have the familiar fourfold possible outcome: Correctly or incorrectly *accepting* an item, correctly or incorrectly *rejecting* it.

Metrologists call the probability of incorrectly accepting an item the “consumer’s risk” and of incorrectly rejecting it the “producer’s risk.” Among other factors, these probabilities depend on the decision rule used for acceptance. Though official documents and scholarly papers typically suggest choosing rules considering the consequences of error for different parties, in practice adherence to this advice varies across subdisciplines, with some using relatively ad hoc conventional choices (Joint Committee for Guides in Metrology (JCGM) 2012; Källgren et al. 2003; Lira 2002). Rules are commonly expressed with an “acceptance interval”: The range within which estimated values must fall for declaring conformance. At one extreme we find the “shared risk” rule: Here the acceptance interval is the whole tolerance interval. This means that as long as an object’s estimated value falls within the tolerance interval—even if it is very close to the tolerance limit—the object is accepted. It is called “shared” because measurement estimations are commonly modeled with a symmetric and unimodal probability density function and, in this case, for estimated values near the end of a tolerance interval, the probabilities of accepting incorrectly and rejecting incorrectly approximate 50%.

In other contexts, the rule involves a *narrower* acceptance interval, thereby reallocating a larger share of risk to the producer’s side. Metrologist Ignacio Lira (2002) gives an interesting rationale:

In practice, the consumer’s risk is usually deemed to be more harmful than the producer’s risk. This is because whereas the costs of rejecting an acceptable item are assumed only by the manufacturer, the decision to accept a faulty product will have deleterious consequences not only on its users, but also on the perceived reliability of the producer. (110)

Lira is emphasizing not so much the short-term interests (or “utilities”) of the users versus producers, but also the perceived reliability (and perhaps trust) that users may have in the producer’s quality control system in place, something with potentially more significant longer-term effects for the overall network of relations.

Consequences of error and uncertainty are familiar topics to the values-in-science literature under the name of inductive risk (see Elliott 2026). Indeed, Rudner’s (1953) original two examples (the level of toxicity in a drug, the quality of buckle belts) could have been taken from a metrology textbook. Yet, engaging more closely with the metrological literature and practice might be fruitful for these philosophical debates.

For example, it's common to frame inductive risk in terms of scientists' "personal probabilities." This framing triggers questions about scientists' cognitive attitudes and whether scientists can be said to "accept" probabilities statements or rather "just have" (beliefs representable by) probabilities that merely await being reported (e.g., [Steel 2015](#)). Although pertinent in many set-ups, this (psychological) perspective seems less relevant in the metrological context, where the method for formally evaluating the uncertainty of measurement outcomes is highly regimented by international bodies. *If* there is not much space here to say that metrologists "choose" probabilities (à la Rudner), it is *not* because they "just have them," but because coming up with an error bar is an institutionally constrained (modeling) activity. The point is not that metrologists make no choices for calculating the uncertainty of measurement results and, hence, are immune to second-order inductive risk considerations. As metrologists acknowledge (e.g., [Lira 2002](#)) and Eran Tal's model-based account helpfully articulates, uncertainty evaluations are model-relative. And modeling choices must be made. But the context in which metrologists make these choices is more institutionally constrained than some discussions of the argument of inductive risk sometimes assume. How this bears on the inductive risk debate is worth studying (see [Wilholt 2009](#)).

More generally, the field of legal metrology—which depends on the collaboration of scientists and regulators from across the globe—might offer important insights for philosophers studying the science/politics interface. Consider again Lira's suggested rationale for reallocating risks. Whether it can be fruitfully connected to the philosophical literature on trust in science and to the arguments for keeping high evidential standards are interesting questions.

34.3.4 *The consequences of being measured*

Unlike samples of water, humans are not indifferent to being measured, as Ian Hacking insisted. Accordingly, it pays to see the act of measurement not only as one of passively gathering information but also as an *intervention* on those being measured ([Rodriguez Duque, Tal, and Barbic 2024](#)) or as a *communicative act* between researchers/clinicians and patients ([McClimans 2024](#)). To illustrate what is easily missed when one takes the gathering-information-perspective only, consider [Rodriguez Duque, Tal, and Barbic's \(2024\)](#) description of some impacts of being measured in mental-health contexts:

[Q]uestionnaires transmit beliefs and values to the persons being measured. The wording and choice of questions inform respondents about what service providers (or other experts) value, and about how experts view the content of the construct. These views may influence or clash with the views of respondents. Questions may be biased towards specific gender, age, race, ethnicity, or socio-economic groups, and thus be perceived as offensive or irrelevant by members of other groups. The absence of certain questions can also impact respondents: if a screening questionnaire for depression does not ask about social media use, a respondent who believes the long time they spend on social media aggravates their mental health may come to doubt themselves or distrust their healthcare provider. ... Finally, the experience of being measured or classified can be harmful in that it could lead respondents to mis-conceptualize or reify their own mental traits ('I answered all questions with 'most of the time', therefore I must be irreparably mentally ill'). (4)

In her in-depth study of patient-reported outcome measures (PROMs), Leah [McClimans \(2024\)](#) takes a communicative perspective. She argues that the philosophical literature "misunderstands the ethical significance" of PROMs (7) by assuming they measure well-being and, accordingly, emphasizing that using the right account of well-being to design these measures should have the desired

consequence of improving overall well-being. McClimans contrasts this *beneficence model* to an *autonomy model*: One that conceives PROMs as “defined by their revolutionary promise to put [patient] perspectives at the center of medical assessment and evaluation” (8). When acts of measurement treat patients as mere sources of information (versus clinicians “primary partners”), claims McClimans, patients have their subjectivity undermined and silenced. Seeing patients as epistemic agents—in genuine dialogue with researchers—requires involving patients much earlier than at the point of asking questions.

34.3.5 Justice-based analyses

An emerging literature grapples with the *politically* profound consequences of some measurement practices by assessing them from the perspective of (epistemic) *justice*.

We have already mentioned the enormous role that metrology plays globally in science and trade. Similarly, industrialized societies have seen the expansion of standardized educational measurement, also leading to large-scale national and international institutions (e.g., the US’s Education Testing Service (ETS), the OECD’s Programme for International Student Assessment (PISA)). In this context, the picture of the lonely scientist exploring how to measure tentative concepts and making small-scale choices (e.g., about a specific measurement model, a specific wording of a question) with downstream effects that raise normative concerns misses something important—the existence of large-scale, institutionalized “measurement systems” or “infrastructures.” Given their tremendous impact on society, argues Zachary [Stein \(2016\)](#), institutionalized measurement systems are part of society’s “basic structure” (in John Rawls’ sense); hence they should be assessed from the perspective of *justice*. When so much of everyone’s educational and job-related life prospects depends on institutionalized educational measurement systems, criteria like objectivity, efficiency, and overall beneficence are all important considerations, but they should form part of a more comprehensive rights-based approach that secures a *just measurement system*, the idea goes.

Similarly, the political significance of specific measurements has also triggered concerns of epistemic justice. For example, not measuring climate science parameters that are essential for adaptation and mitigation purposes in some highly vulnerable areas of the globe (Elabbar [forthcoming](#)) or social science indicators (e.g., cost of living) designed from some marginalized standpoints ([Thoma 2024](#)) can disadvantage people living in those areas or having those standpoints—they may lack the epistemic resources to assess their own predicament and/or to express their predicament persuasively in public settings. This can affect the quality of public deliberation, recognition, accountability, compensation, and so on (Elabbar [forthcoming](#); [Thoma 2024](#)). In such cases, the argument goes, the politically significant role these measurements play requires going beyond beneficence considerations—which commonly dominate discussions of science’s research agenda (e.g., [Kitcher 2001](#))—and taking an *epistemic justice* perspective.

34.3.6 Values and validity

To promote links between the values-in-science and the measurement literatures, I close this section illustrating how an influential tradition of psychological and educational researchers grapples with values in their measurement practices. Within this tradition, which has shaped the latest version of the highly influential *Standards of Educational and Psychological Testing* ([AERA, APA, and NCME 2014](#)), a central claim is that the quality of measurements (their “validity”) should be assessed not only in evidential terms but also in terms of their socially significant *consequences*. A prominent figure here is Samuel Messick. “If, as an intrinsic part of the overall validation process, we weigh the

actual and potential consequences of our testing practices in light of considerations of what future society might need or desire,” says Messick, “then test validity comes to be based on ethical as well as evidential grounds” (1980, 1020). (A question worth pursuing is how related Messick’s approach is to the “coupled ethical-epistemic analysis” advocated by (e.g.) [Katikireddi and Valles \(2015\)](#) in a public health context.)

For Messick, these ethical considerations are not restricted to the *use* of measurements: “the measurement of [psychological and educational] characteristics entails value judgments—at all levels of test construction, analysis, interpretation, and [not only] use” (1980, 1013). When it comes to interpretation, for example, [Messick \(1980\)](#) highlights the social *connotations* that different interpretations of measurements can have, each with resulting social consequences:

a measure interpreted in terms of “flexibility versus rigidity” would be utilized differently if it were instead labeled “confusion versus control.” ... So would a variable like ‘stress’ if it were relabeled “challenge.” The point is not that we would make a concept like stress into a good thing by renaming it but that by not presuming it to be a bad thing we would investigate broader consequences, facilitative as well as debilitative. (1022)

This passage has interesting affinities to [Anderson’s \(2004\)](#) well-known discussion of framing and conceptualization in divorce research, as well as with the discussion of these themes by [Elliott \(2017, chap.7\)](#), where the examples are not restricted to the social sciences.

Now, when it comes to the consequences of *use*—selection of students and workers—Messick is adamant that unintended consequences be central in validity assessments. Careful scrutiny should be given to whether a particular social group is disadvantaged by a particular test. There is now a vast psychometric literature offering formal accounts of what bias/unfairness in tests amounts to (e.g., [Borsboom, Romeijn, and Wicherts 2008](#)).

Someone may claim that socially significant consequences of *use* are irrelevant when assessing the values behind measurements: We should distinguish between the measurements *themselves* and our uses of them. Accordingly, we shouldn’t infer that the measurements themselves are value-laden from the claim that (the likely) uses of the measurements raise normative concerns (see discussion in [Schroeder \(2017\)](#)). Of course, Messick argued that the consequences to be assessed are not only those of *using* the measurement but also those resulting from the interpretations given to the measurements. That aside, as I interpret it, part of the argument made by this tradition is that the distinction between measurements “themselves” and their uses is not at all appealing in the present context—it gets the scope of analysis wrong. If the measurement practice at stake is largely meant to be the core (if not definitive) input for a decision, as it is in many educational and psychological contexts, little is gained by excluding from the assessment of a measurement practice the assessment of the consequences of the application in view, the idea goes. As [Kane \(1992, 530\)](#) puts it, “Most tests are ... linked to some decision. If the test scores were not relevant to any decision, it is not clear why the test would be given.”

34.4 The content of measurands

Finally, some influential arguments focus on the *content* of the concept intended to be measured (the “measurand”) and on how this content is determined.

The practice of measurement—of gathering empirical information about given measurands—assumes some things about these measurands such that it is possible to claim to have measured *them* (versus something else). It assumes that measurands can be characterized rather precisely so that, e.g., it is clear what does—and does not—fall under it (if it is a categorical concept), what “amounts” of it are (if it is a quantitative concept), which disturbing factors must the measurement procedure

control for, etc. After all, if the characterization is vague, when assessing different candidate measurement methods, one quickly runs out of grounds for adjudicating between them. If they deliver (very) different results, we've got a (serious) problem. Accordingly, we must be *specific* on what the measurand is. Now, if there is leeway on how to specify the concept, values may enter, the idea goes.

Consider a seemingly straightforward concept within macroeconomics: *The unemployment rate*. Roughly, it's the share of the economically active population (e.g., excluding children, students) that is currently not working and actively searching for a job. Although detailed, this characterization is not detailed enough. What counts as "actively searching" (how active, for how long)? or as "not working" (one day a week, part time)? Given the role that unemployment plays in social policy (e.g., determining who gets unemployment insurance, or pension benefits), different precisifications may have different policy repercussions for different social groups. Different precisifications may also make more/less sense depending on the assumed questions that the unemployment rate is meant to be informing (Anderson 1995, 45–6): For example, "actively searching" can be specified strictly, so that those who want a job but stopped searching because they just cannot find one ("discouraged workers") are excluded, or loosely, so as to include discouraged workers. The former makes more sense if the goal is to measure a macroeconomic factor that (background theory says) heavily influences wages. All else equal, however, this choice may lead to a more austere unemployment insurance scheme and, more generally, it may be taken to express a narrower conception of the welfare duty of the state. The alternative option arguably bears a closer connection to *other* macroeconomic variables of interest like poverty (as well as further social phenomena such as crime), and it may lead to a more generous insurance scheme, expressing a stronger welfare duty.

If we want to adjudicate between these (and other) precisified versions of the starting characterization, we seem to need further considerations. Perhaps we must consider what the *purposes or aims* behind the practice are (e.g., Anderson 1995; Intemann 2015). Perhaps we must consider the *consequences* of the different choices (Elliott 2017). (One can read claims about purposes as claims about which consequences are more salient, so there need not be any conflict here.) Claims about purposes (or relevant consequences) depend on judgments of what is more/less socially significant (Anderson 1995), that is, on value-judgments.

Consider now the scientific study of *well-being*. Roughly, well-being is what makes someone's life go well. Evidently, this characterization cannot guide measurement practices much. How should well-being be precisified for purposes of measurement? Well-being is a quintessential evaluative concept: It plays a central role in moral discourse—e.g., according to influential moral theories, considerations of aggregated well-being *clinch* the answer to all moral questions. Hence, Alexandrova (2018) argues, precisifying well-being requires presupposing a normative judgment about the *nature* of this concept. To precisify well-being, for example, as a favorable balance of positive over negative emotions requires committing to a hedonist view about what makes life go well, which is just one morally controversial position among others.

The argument that calls for purposes/consequences for choosing precisifications (as in the unemployment example) is typically lumped together with Alexandrova's, e.g., as instances of the "Conceptual Argument" (Elliott 2026). Though reasonable, there are noteworthy differences (see also Elliott 2022, 32–3): The former argument makes no claim about the *evaluative nature* of the concept whereas the latter makes no reference to *purposes/consequences*. Moreover, the former argument is typically accompanied by the claim that the concept at stake is (not sharp but) *ballung*: The objects falling under the concept don't share a set of necessary and sufficient conditions, or of causes and effects, but rather a family resemblance. For instance, "the different violent conflicts that get called civil wars all differ from one another" in their internal characteristics as well as in their causes and effects (Cartwright and Runhardt 2014, 268). Accordingly, there are many ways of precisifying "civil war," each of them emphasizing some aspects (or causal relations) while downplaying others.

In contrast, Alexandrova’s argument doesn’t involve a claim about the heterogeneity of the cases falling under the concept, but one about the “nature” of the concept: It being evaluative, or more specifically, a *thick concept*, i.e., one that both describes and evaluates. Values enter because we must take a moral stance about what (e.g.) well-being *is* to defend a particular precisification of it.

Put thus, the two arguments indeed look different—one ties precisification of heterogeneous concepts to purposes/consequences, the other ties precisification to normative commitments about the ethical concept at stake. But are they completely independent? Sometimes both arguments are run for the same concept (for biodiversity, see [Habib \(2015\)](#); [Intemann \(2015\)](#), respectively), and sometimes by the same authors (both [Reiss \(2018\)](#) and [Thoma \(2024\)](#) discuss inflation (or “cost of living”) as a case where both arguments apply). More conceptual work is needed here (e.g., [Lewens 2025](#)).

A closing thought: Central to most arguments for values-in-science is the assumption that scientists face contingent choices (e.g., [Brown 2020](#); [Elliott 2026](#)). This is also true for the arguments just reviewed: When precisifying for purposes of measurement, scientists have leeway to choose. But how much leeway do scientists have, really?

Let’s focus on the quantitative measurement of theoretical (unobservable) concepts. Scientists having much leeway presuppose a relative abundance of epistemically defensible candidate quantifications, such that scientists can choose the morally defensible ones among the epistemically defensible ones. Yet, history teaches, the quantification of unobservable concepts like temperature ([Chang 2004](#); [Sherry 2011](#)) or earthquake size ([Howell 1990](#); [Miyake 2017](#)) is anything but straightforward: Arriving at quantitative concepts that are well-coordinated with empirical procedures is hard. What if, then, on a closer look, there isn’t that much freedom to choose? Ohnesorge (2025) articulates this point, charging the literature for assuming without argument that quantitative measurement can be achieved in various ways, such that we can choose the normatively better ways among the epistemically successful ones.

Consider [Schroeder’s \(2017\)](#) discussion of population summary measurements: Quality-adjusted life years (QALY) and disability-adjusted life years (DALY). These measures combine (multiply) *years* of life with a *quality* factor that is meant to capture how healthy/unhealthy the person is. Schroeder argues that several choices for specifying QALYs and DALYs are value-laden. To name two, researchers used normative considerations when deciding which weights to give to people of different ages (e.g., “some age groups are more important for society’s flourishing, hence their gains/losses deserve more weight”) and what “discount factor” to use to make present and future years of life lost comparable (e.g., “how much should the current generation compromise for future generations?”).

Why some cases of quantitative measurement offer scientists much leeway while others much less so? Perhaps in some of the former—especially when the measures combine rather different indicators—researchers may be after a different notion of measurement altogether: Researchers may be best understood as aiming *not* to *discover* a quantitative concept that helps in explaining/predicting phenomena, but to *value* some state of affairs ([Hausman 2015](#); [Larroulet Philippi 2023](#), Unpublished Manuscript). The latter requires researchers to decide (say) about weights and discount factors on evaluative grounds, not to find out whether they can coordinate candidate concepts and procedures coherently.

Think of “earthquake size.” A seismology textbook starts its chapter on measurement thus: “The size of an earthquake can be measured in many ways. The number of deaths and injuries and the value of property damaged and destroyed give measures of the seriousness of the event in human terms” ([Howell 1990](#), 97). If we were to combine deaths, injuries, and property damage into an overall measure of earthquake size, we would also face many choices—do all deaths count equal? How many injuries, of what seriousness, count as one death? Etc.

What would be the measurand of this exercise? Not “earthquake size” in a physical sense (e.g., how much energy the earthquake released), but arguably something like *the disvalue* that humans

attribute to earthquakes.³ This explains why our measurement procedure doesn't control for distance to the epicenter, construction quality, and other details that affect how bad earthquakes are but confound how much energy was released. It also explains why we combine these different ("incommensurable") things (deaths, injuries, property damage) by judging their relative seriousness. The vast number of choices to make, and the need for normative reasons to settle those choices, makes perfect sense if we are engaging in an all-things-considered *valuation* exercise.

My suggestion, then, is that some cases of measurement with few empirical constraints and much leeway may be best seen as exercises in valuation (versus in the discovery of unidimensional quantitative concepts for purposes of explanation/prediction). Even if values influence discovery, it seems wise to distinguish such (subtler) value-influences from those straightforwardly at play in *valuing* exercises.

34.5 Conclusion

This chapter reviews influential arguments for the claim that values play a role in measurement practices. Some arguments concern background assumptions behind the pursuit of measurement, others the various consequences of measurement practices (or aspects of them), and others how scientists specify the content of measurands. I hope the chapter demonstrated the importance for the values-in-science literature to engage with the historical-philosophical literature on measurement.

Acknowledgments

Thanks to Alessandra Basso, Marion Boulicault, Ahmad Elabbar, Kevin Elliott, Jacob Stegenga, Miguel Ohnesorge, and Anna Alexandrova's PhD group for helpful comments and suggestions.

Notes

- 1 To be sure, quantification efforts need not be explained only by arguments to the effect that quantifying some phenomenon "makes sense." As historian Theodore Porter (1995) argued, quantification can sometimes be a strategic solution to a political problem—that of bureaucrats' political weakness. Using numbers to assess (say) "intelligence" or "scholastic aptitude" may enable weak educational officials to elude criticism targeting their decisions ("blame the numbers, not me!").
- 2 Thanks to Eran Tal for pointing me toward this literature.
- 3 Hausman (2015) makes the analogous claim about measures of overall health.

References

- AERA, APA, and NCME. 2014. *Standards for Educational and Psychological Testing*. American Psychological Association.
- Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *British Journal for the Philosophy of Science* 69: 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Anderson, Elizabeth. 1995. "Knowledge, Human Interests, and Objectivity in Feminist Epistemology." *Philosophical Topics* 23: 27–58. <https://www.jstor.org/stable/43154207>.
- Anderson, Elizabeth. 2002. "Situated Knowledge and the Interplay of Value Judgments and Evidence in Scientific Inquiry." In *In the Scope of Logic, Methodology and Philosophy of Science*, edited by Peter Gärdenfors, Jan Wolenski, and Katarzyna Kijania-Placek. Kluwer. https://doi.org/10.1007/978-94-017-0475-5_8.
- Anderson, Elizabeth. 2004. "Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19: 1–24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>.
- Berlin, Isaiah. 2013. *The Roots of Romanticism*. Princeton University Press.
- Borsboom, Denny, Jan-Willem Romeijn, and Jelte Wicherts. 2008. "Measurement Invariance Versus Selection Invariance: Is Fair Selection Possible?" *Psychological Methods* 13: 75–98. <https://psycnet.apa.org/doi/10.1037/1082-989X.13.2.75>.

- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Cartwright, Nancy, and Rosa Runhardt. 2014. "Measurement." In *Philosophy of Social Science: A New Introduction*, edited by Nancy Cartwright and Eleonora Montuschi. Oxford University Press.
- Chang, Hasok. 2004. *Inventing Temperature*. Oxford University Press.
- Coen, Deborah. 2013. *The Earthquake Observers*. University of Chicago Press.
- Daston, Lorraine. 1996. Review of *The Values of Precision*, by Norton Wise. *Isis* 87: 517–88.
- Elabbar, Ahmad. forthcoming. "Varying Evidential Standards as a Matter of Justice." *British Journal for the Philosophy of Science*. <https://doi.org/10.1086/727429>.
- Elliott, Kevin. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin. 2026. "Arguments against the Value-Free Ideal." In *The Routledge Handbook of Values and Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Elliott, Kevin C., and Rebecca Korf. 2024. "Values in Science: What Are Values, Anyway." *European Journal for Philosophy of Science* 14: 1–24. <https://doi.org/10.1007/s13194-024-00615-3>.
- Habib, Allen. 2015. "Biodiversity and Values in Science." *Ethics, Policy & Environment* 18: 30–33. <https://doi.org/10.1080/21550085.2015.1016962>.
- Hausman, Daniel. 2015. *Valuing Health*. Oxford University Press.
- Howell, Benjamin. 1990. *An Introduction to Seismological Research*. Cambridge University Press.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5: 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Joint Committee for Guides in Metrology (JCGM). 2012. *Evaluation of Measurement Data—The Role of Measurement Uncertainty in Conformity Assessment*. www.bipm.org/utls/common/documents/jcgm/JCGM_106_2012_E.pdf.
- Källgren, Håkan, Margreet Lauwaars, Bertil Magnusson, Leslie Pendrill, and Phillip Taylor. 2003. "Role of Measurement Uncertainty in Conformity Assessment in Legal Metrology and Trade." *Accreditation and Quality Assurance* 8: 541–7. <https://doi.org/10.1007/s00769-003-0707-8>.
- Kane, Michael. 1992. "An Argument-Based Approach to Validity." *Psychological Bulletin* 112: 527–35.
- Katikireddi, Srinivasa V., and Sean A. Valles. 2015. "Coupled Ethical-Epistemic Analysis of Public Health Research and Practice: Categorizing Variables to Improve Population Health and Equity." *American Journal of Public Health* 105: e36–42. <https://doi.org/10.2105/AJPH.2014.302279>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Koyré, Alexandre. 1953. "An Experiment in Measurement." *Proceedings of the American Philosophical Society* 97: 222–237. <https://www.jstor.org/stable/3143896>.
- Larroulet Philippi, Cristian. 2023. "On the Challenges of Measurement in the Human Sciences." PhD diss., University of Cambridge. <https://doi.org/10.17863/CAM.102194>.
- Larroulet Philippi, Cristian. Unpublished Manuscript. "Quantifying the Human: Values in Measurement or Measuring Value?"
- Lewens, Tim. 2025. "Towards a Direct Role for Values in the Heart of Science." *Philosophy of Science* 1–25. <https://doi.org/10.1017/psa.2025.18>.
- Lira, Ignacio. 2002. *Evaluating the Measurement Uncertainty: Fundamentals and Practical Guidance*. Institute of Physics Publishing.
- McClimans, Leah. 2024. *Patient-Centered Measurement*. Oxford University Press.
- Messick, Samuel. 1980. "Test Validity and the Ethics of Assessment." *American Psychologist* 35: 1012–27. <https://psycnet.apa.org/doi/10.1037/0003-066X.35.11.1012>.
- Michell, Joel. 1999. *Measurement in Psychology*. Cambridge University Press.
- Miyake, Teru. 2017. "Magnitude, Moment, and Measurement: The Seismic Mechanism Controversy and Its Resolution." *Studies in History and Philosophy of Science* 65: 112–20. <https://doi.org/10.1016/j.shpsa.2017.02.002>.
- O'Connell, Joseph. 1993. "Metrology: The Creation of Universality by the Circulation of Particulars." *Social Studies of Science* 23: 129–73. <https://doi.org/10.1177/030631293023001005>.
- Ohnesorge, Miguel. 2025. "Should We Align Quantitative Measures with Stakeholder Values?" *Philosophy of Science*. <https://doi.org/10.1017/psa.2025.10154>
- Planck, Max. 1949. "The Meaning and Limits of Exact Science." *Science* 110: 319–27. <https://doi.org/10.1126/science.110.2857.319>.
- Porter, Theodore. 1995. *Trust in Numbers*. Princeton University Press.
- Reiss, Julian. 2018. "Measurements and Value Judgements." In *The Routledge Handbook of the Philosophy of Economics*, edited by Julian Reiss and Conrad Heilmann. Routledge.

- Rodriguez Duque, Sebastian, Eran Tal, and Skye Pamela Barbic. 2024. "The Role of Ethical and Social Values in Psychosocial Measurement." *Measurement* 225: 113993. <https://doi.org/10.1016/j.measurement.2023.113993>.
- Rose, Steven, Richard Lewontin, and Leon Kamin. 1984. *Not in Our Genes*. Penguin Books.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgements." *Philosophy of Science* 20: 1–6. <https://doi.org/10.1086/287231>.
- Schaffer, Simon. 1988. "Astronomers Mark Time: Discipline and the Personal Equation." *Science in Context* 2: 115–45. <https://doi.org/10.1017/S026988970000051X>.
- Schroeder, S. Andrew. 2017. "Value Choices in Summary Measures of Population Health." *Public Health Ethics* 10: 176–87. <https://doi.org/10.1093/phe/phw032>.
- Sherry, David. 2011. "Thermoscopes, Thermometers, and the Foundations of Measurement." *Studies in History and Philosophy of Science* 42: 509–24. <https://doi.org/10.1016/j.shpsa.2011.07.001>.
- Steel, Daniel. 2015. "Acceptance, Values, and Probability." *Studies in History and Philosophy of Science* 53: 81–8. <https://doi.org/10.1016/j.shpsa.2015.05.010>.
- Stein, Zachary. 2016. *Social Justice and Educational Measurement*. Routledge.
- Tal, Eran. 2016. "Making Time: A Study in the Epistemology of Measurement." *British Journal for the Philosophy of Science* 67: 297–335. <https://doi.org/10.1093/bjps/axu037>.
- Thoma, Johanna. 2024. "Social Science, Policy and Democracy." *Philosophy and Public Affairs* 52: 5–41. <https://doi.org/10.1111/papa.12250>.
- Ward, Zina. 2021. "On Value-Laden Science." *Studies in History and Philosophy of Science* 85: 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Wijsen, Lisa D., Denny Borsboom, and Anna Alexandrova. 2022. "Values in Psychometrics." *Perspectives on Psychological Science* 17: 788–804. <https://doi.org/10.1177/17456916211014183>.
- Willholt, Torsten. 2009. "Bias and Values in Scientific Research." *Studies in History and Philosophy of Science* 40: 92–101. <https://doi.org/10.1016/j.shpsa.2008.12.005>.
- Wise, M. Norton, ed. 1995. *The Values of Precision*. Princeton University Press.
- Wise, M. Norton, and Crosbie Smith. 1986. "Measurement, Work and Industry in Lord Kelvin's Britain." *Historical Studies in the Physical and Biological Sciences* 17: 147–73. <https://doi.org/10.2307/27757577>.

35

SCIENCE, VALUES, AND RACE CORRECTION

Zinhle Mncube

35.1 Introduction

Race correction, or the use of race as a predictor in clinical algorithms to estimate health risks, is highly contentious in the US. In this chapter, I argue that, descriptively, at least part of the debate about whether to include or exclude a race adjustment in clinical predictive algorithms can be plausibly understood as a classic case of inductive risk. In [Section 35.2](#), I explain the practice of race correction using the case study of the CKD-EPI_{cr} equation used to predict kidney function. In [Section 35.3](#), I introduce eliminativists and conservationists as the two sides of the race correction debate. In [Section 35.4](#), I outline [Rudner \(1953\)](#) and [Douglas's \(2000\)](#) classic arguments from inductive risk. In [Sections 35.5 and 35.6](#), I show that the argument from inductive risk disambiguates some of eliminativists and conservationists' competing claims about using race in the CKD-EPI_{cr} equation. My analysis adds to the values in science literature by showing that values have plausibly played a role in the currently more popular stance in the race correction debate to reject the use of race in the construction of clinical algorithms.

35.2 What is race correction?

Race correction refers to the process of systematically adjusting the outputs of clinical predictive algorithms, risk predictor tools, and practice guidelines according to a patient's race or ethnicity. Today, race correction is largely and controversially used in several predictive algorithms in the US in clinical specializations like nephrology, cardiology, endocrinology, and obstetrics, amongst others ([Vyas, Eisenstein, and Jones 2020](#)). I refer coextensively to the use of race or ethnicity as a predictor in a clinical predictive algorithm as “race correction,” a “race adjustment,” or a “race-adjusted” algorithm. Race-adjusted clinical algorithms incorporate race alongside other predictor variables to estimate the probability that a health outcome is present or will occur for a patient over a period. These algorithms are used in clinical practice to individualize medical risk assessment and inform clinical decision making. Consider a prominent example of a race-adjusted clinical algorithm: The eGFR_{cr} equation.

Accurate assessment of kidney function is crucial for clinical decisions such as the referral of patients to kidney specialists, and eligibility for the kidney transplant waitlist. Glomerular filtration rate (GFR)—the flow of filtered fluid through the kidneys—is considered the best measure of functional

kidney capacity. The most convenient way of estimating GFR in clinical practice is through equations (known as “eGFRcr equations”) that include clearance of the protein “creatinine” from the blood. The most widely used eGFRcr equation in the US is the 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI_{cr}) version (henceforth, the “CKD-EPI_{cr} equation” or “2009 CKD-EPI_{cr} equation”) (Levey, Stevens et al. 2009). This equation estimates GFR (known as “eGFR”) based on a patient’s age, sex, serum creatinine, and *race* (Black or non-Black). The effect of the race adjustment is that it multiplies the eGFR for all patients identified as Black by a factor of 1.16 more than for non-Black patients with similar clinical characteristics.

This upward inflation of eGFR has significant clinical consequences for Black patients. Suppose that Enzo is a 50-year-old woman with a serum creatinine level of 2 mg/dL. The threshold for referral to a kidney specialist is an eGFR of less than 30 mL/min/1.73 m². If Enzo is identified as White, her race-adjusted eGFR is 28 mL/min/1.73 m², and she will be referred to a specialist. If she is identified as Black, her race-adjusted eGFR is 33 mL/min/1.73 m², and she will likely *not* be referred to a specialist. The equation’s developers, Levey, Stevens et al. (2009), explain the need for the race adjustment as due to observed higher serum creatinine levels among their Black study participants, compared to their non-Black ones—their so-called muscle mass hypothesis (Roberts 2021). For Levey and colleagues, it is justified to include a race adjustment in part because the purported causal mechanism by which race affects serum creatinine-based differences in kidney function is muscle mass. This practice of race correction has recently come under heavy scrutiny.

35.3 The race correction debate

What I call “the race correction debate” is about whether we should retain or eliminate the use of race as a predictor in clinical predictive algorithms. This is a philosophical debate about the scientific value of race as a predictor in these algorithms. What do I mean by the “scientific value” of race? The race correction debate is about whether race *hinders* or *helps* the prediction of individualized medical risk in clinical practice. Does using race improve the predictive accuracy of clinical algorithms? Can clinicians use race consistently as a variable, such that predictions of risk are reliable not just for some patients but across the population? Is there a solid theoretical and evidential basis for the use of race in race-adjusted algorithms? Does using race in these algorithms entail favorable consequences for patients, and modern medicine more generally? In this chapter, I focus on the dispute about the race-adjusted 2009 CKD-EPI_{cr} equation as an illustrative case of arguments in the race correction debate.

Critics of race correction defend a position I call “eliminativism.” Eliminativists argue that we should immediately remove race from clinical predictive algorithms. The eliminativist position is composed of metaphysical, epistemological, and ethical arguments. Metaphysically, most eliminativists maintain that including race in clinical algorithms presupposes that “race” is biologically real. However, they argue that we should cease race correction because race is a social construct (Braun et al. 2021; Cerdeña, Plaisime, and Tsai 2020; Vyas, Eisenstein, and Jones 2020). Epistemologically, eliminativists question the scientific veracity of equation developers’ rationale for including race in clinical algorithms. For example, critics claim that we should remove race from the CKD-EPI_{cr} equation because Levey, Stevens et al.’s (2009) muscle mass hypothesis is scientifically dubious (Eneanya, Boulware, and Tsai 2022; Marzinke and Greene 2022; Roberts 2021).

Ethically, eliminativists cite some empirical studies that compare the performance of the CKD-EPI_{cr} equation with and without race to claim that the upward adjustment of eGFR for Black patients may exacerbate systemic health inequities in the US (e.g., Eneanya, Boulware, and Tsai 2022, 90; Marzinke and Greene 2022, 11; Mohottige et al. 2022, 317). Specifically, for them, using race risks erroneously inflating Black patients’ eGFR, thereby leading to unjust delays in the

treatment of their CKD (Delgado et al. 2021; Eneanya, Boulware, and Tsai 2022). This is why, for these critics, eliminating race from clinical algorithms is a step toward mitigating health inequities (see Cerdeña, Plaisime, and Tsai 2020). Today, the nationally convened National Kidney Foundation and American Society of Nephrology (NKF-ASN) Task Force that was formed to reassess the use of race in the 2009 CKD-EPI_{cr} equation recommends that medical institutions in the US use the race-free version (Delgado et al. 2021). Eliminativism is undoubtedly the popular position in the race correction debate.

Proponents of race correction defend a position I call “conservationism.” Conservationists accept that it may be difficult to reliably measure patients’ race, but they insist that we should retain the use of race in clinical algorithms until better predictors of health outcomes become available (Levey, Titan et al. 2020; Velasco and Snodgrass 2021, 828). Conservationists claim that the race correction in the CKD-EPI_{cr} equation is justified because it improves the accuracy of GFR assessment (Levey, Tighiouart et al. 2020; Manski 2022; Manski et al. 2023; Velasco and Snodgrass 2021; Williams 2023). Conservationists warn that “arbitrarily” eliminating race adjustments from clinical algorithms will foreseeably harm all patients (Manski 2022; Velasco and Snodgrass 2021). For them, race-free clinical risk assessments would be inaccurate. Inaccurate risk assessments would then harm patients because they would receive sub-optimal clinical care. For some conservationists, using race as a predictor optimizes clinical decisions and maximizes patient well-being (Manski 2022, 2111).

How can using race in the CKD-EPI_{cr} equation be both unjustified for eliminativists, and yet justified for conservationists? How can the race adjustment risk exacerbating health inequities for eliminativists, and yet maximize patient well-being for conservationists? My thesis in this chapter is that descriptively it is plausible to interpret at least part of what is at issue in the race correction debate as a classic case of inductive risk, and doing so helps us disambiguate some competing eliminativist and conservationist claims.

35.4 The argument from inductive risk

“Inductive risk” refers to the risk of erroneously accepting or rejecting a hypothesis based on a body of evidence. The formulation of the classic argument for inductive risk is based on Richard Rudner analysis.¹ Rudner (1953) argues for a legitimate and inescapable role of non-epistemic values in scientists’ inductive inference from the evidence to accepting or rejecting a hypothesis. He claims that since a body of evidence cannot confirm a scientific hypothesis with complete certainty, there is always inductive risk—the risk that scientists mistakenly accept a false hypothesis (i.e., a false positive) or reject a true one (i.e., a false negative). To balance the risk of each error, Rudner claims that scientists must decide on an evidential standard that warrants accepting the hypothesis. Sometimes serious non-epistemic consequences follow from mistakenly accepting or rejecting a scientific hypothesis. In these cases, Rudner (1953, 2) insists that whether scientists decide to demand a relatively high or low degree of evidential support before accepting the hypothesis depends on the seriousness of the non-epistemic consequences that might result from an error in either direction. For him, judgments of the severity of inductive risk, and thereby scientific evidentiary standards, are value-laden. Thus, it is in accepting or rejecting hypotheses that scientists inevitably make ethical value judgments (Rudner 1953).

Heather Douglas famously accepts and extends Rudner’s argument to claim that scientists balance inductive risks using non-epistemic values not just in the final hypothesis acceptance step but *throughout* the scientific research process including in “the choice of methodology, gathering and characterization of the data, and interpretation of the data” (Douglas 2000, 565). She illustrates her argument by analyzing laboratory studies on the carcinogenicity of dioxins on animals. One of her examples involves the methodological choice of a statistical significance level.

It is necessary to choose a statistical significance level for conducting the laboratory animal studies to decide whether a particular toxicological response in exposed animals has been caused by dosing of dioxins. Douglas (2000, 566) argues that in setting a statistical significance level for the studies, scientists must decide on an appropriate balance between false positive and false negative errors. That is, the chance of wrongly accepting that dioxins cause a toxicological response, when they don't; or rejecting that dioxins cause a toxicological response, when they in fact do.

She maintains that setting a relatively high statistical significance level for the studies communicates a tolerance of false negatives (Douglas 2000, 567). For Douglas, this choice would likely have the policy implication of underregulation of dioxins, as these chemicals would not appear as toxic as they are. Underregulation of harmful dioxins could foreseeably damage public health. She argues that setting a relatively lower statistical significance level communicates a tolerance of false positives (Douglas 2000, 567). For her, this choice would likely lead to overregulation of dioxins, as then these chemicals would appear more harmful than they are. In this case, the industries that use dioxins would likely incur greater financial costs because of the regulations. Douglas maintains that scientists should set the statistical significance level by trading off the different consequences associated with the false positive and negative results. For her, this risk assessment trade-off involves considering non-epistemic values, and this is how non-epistemic values legitimately shape the methodological choice of a statistical level.

35.5 To include or to not include race?

In what follows, I argue that much like Rudner (1953) and Douglas's (2000) analyses, descriptively, inductive risk can plausibly illuminate some of the debate between eliminativists and conservationists about the race-adjusted CKD-EPI_{cr} equation. There is inductive risk in the methodological choice to exclude or include a race adjustment in the CKD-EPI_{cr} equation because each option has different consequences for clinical practice, clinical decision making, and the lives of patients. While I acknowledge that is difficult to be sure exactly which factors are descriptively responsible for scientists differing conclusions. I will show that *part* of what seems to be of issue between eliminativists and conservationists is whether there is enough evidence to accept or reject this hypothesis *H*:

(*H*) Using a race adjustment in the 2009 CKD-EPI_{cr} equation generates more accurate predictions of GFR for patients to inform kidney-related clinical decisions in clinical practice.

When one includes or excludes a race adjustment in the CKD-EPI_{cr} equation, one requires a relatively higher or lower standard of evidence before accepting hypothesis *H*. Therefore, we can describe this methodological choice as an answer to how much evidence suffices to accept or reject that the race-adjusted CKD-EPI_{cr} equation generates more accurate estimations of GFR for patients (i.e., hypothesis *H*) (Biddle 2016). I will show that in some of their arguments, eliminativists and conservationists seem to *depend* on non-epistemic value judgments about the practical and ethical consequences of wrongly accepting or rejecting *H* (i.e., standard inductive risk considerations). As such, I maintain that descriptively, it appears that, non-epistemic values in part guide the methodological choice to remove or retain race in this example of the race correction debate.

Before I present my argument, it is important to note the ways in which non-epistemic interests and values are *embedded* within hypothesis *H*. First, equation developers, Levey, Stevens et al. (2009), constructed the race-adjusted CKD-EPI_{cr} equation with an epistemic purpose in mind: To accurately *predict* the GFR of patients. But non-epistemic interests and values also underpin this purpose (see Parker and Winsberg 2018, section 2). Clinicians are interested in predicting the GFR for individual patients in clinical practice: This reflects a value for a *practical* way of predicting GFR.

For example, “mGFR” is the gold-standard and most direct way of measuring GFR using clearance of exogenous filtration markers like inulin. Yet mGFR is not used in routine clinical practice because it is too impractical and expensive to measure GFR in the way it requires. From *H*, the GFR estimations are used to inform clinical decision making. This has social implications because the health and lives of patients are directly impacted by (erroneous) predictions of GFR.

Second, [Ward’s \(2021\)](#) taxonomy outlines at least four ways in which epistemic and non-epistemic values could contribute to the scientific choice to accept hypothesis *H*. Hypothesis *H* could be value-laden because: (1) Scientists are consciously motivated by values when deciding to accept it; (2) values could act as justifying reason for accepting it or not; (3) values causally influence whether or not *H* is accepted (they serve as causal effectors); or (4) accepting or rejecting *H* itself promotes or undermines specific values.

Hypothesis *H* also suggests the foreseeable consequences of error to be considered when deciding to remove or retain race in GFR estimation ([Lusk and Elliott 2022](#)). At minimum, these consequences are regulatory (i.e., which equation ends up being used in routine clinical practice), and they relate to clinical decision making, general medical care, the lives of individual patients, and public health (see [Delgado et al. 2021](#), 2999–3007; [Inker et al. 2021](#), 1745).

35.5.1 Underdetermination

To see that we can describe some of the basis of the disagreement between eliminativists and conservationists as how much evidence is sufficient to declare race-adjusted GFR estimations as more accurate (i.e., accept hypothesis *H*), it is important that I first establish the following. As in other cases of inductive risk, the currently available empirical evidence on GFR estimation transiently underdetermines the methodological choice of a race adjustment in the CKD-EPI_{cr} equation. That is, as it stands, the very same body of evidence is compatible with *alternative* methodological choices: That the CKD-EPI_{cr} equation that *includes* a race adjustment leads to more accurate eGFR results for patients; and that the CKD-EPI_{cr} equation that *excludes* it leads to the same. Because there is this gap between the empirical evidence and the methodological choice, depending in part on what an eliminativist or conservationist theorist *values*, they will be convinced by different amounts of evidence, and thus find different methodological options—excluding or including race as a predictor in the CKD-EPI_{cr} equation—as leading to more accurate estimated GFR results.

Based on the currently available empirical evidence, it is arguably uncertain whether including or excluding a race adjustment in the CKD-EPI_{cr} equation leads to more accurate GFR estimations for patients. “Accuracy” is a measure of how closely an equation’s GFR estimations match patients’ mGFR values. An mGFR value is a direct estimate of a patient’s actual GFR, whereas an eGFR value is only an indirect *prediction* of GFR. On one hand, in the study population in which the equation was derived, [Levey, Stevens et al. \(2009\)](#) found that for 8,254 racially diverse Americans, including a race adjustment improved the accuracy of GFR estimation for all participants. Using the same study population, Levey and colleagues ([Levey, Tighiouart et al. 2020](#)) compared the performance of the 2009 CKD-EPI_{cr} equation with and without race. They found that eliminating the race adjustment resulted in a systematic *underestimation* of mGFR, worse so for Black study participants compared to non-Black ones ([Levey, Tighiouart et al. 2020](#)). Similarly, Hsu, Parikh et al. (2021) found that completely excluding race in the CKD-EPI_{cr} equation led to an underestimation of mGFR for Black participants. As such, one could plausibly take this empirical evidence to support that using a race adjustment in the CKD-EPI_{cr} equation leads to more accurate predictions of GFR for patients.

On the other hand, in a study of 1,648 self-identified Black participants, [Zelnick et al. \(2021\)](#) found that including the race adjustment in the CKD-EPI_{cr} equation *overestimated* mGFR for this population. [Inker et al. \(2021\)](#) evaluated the performance of the CKD-EPI_{cr} equation in a pooled

study of over 13,000 racially diverse participants. These authors found that completely eliminating race in the equation (their so-called 2021 CKD-EPI_{cr} equation) improved the accuracy of GFR estimations for Black participants but slightly reduced the accuracy for non-Black participants (Inker et al. 2021, 1744).

The NKF-ASN Task Force compared the performance of 26 approaches to GFR estimation using an external validation dataset of 5,040 study participants (Delgado et al. 2021). Their findings were like Inker et al.'s (2021): The race-adjusted CKD-EPI_{cr} equation overestimated mGFR for Black participants. The race-free 2021 CKD-EPI_{cr} equation had better accuracy for Black participants, but moderate inaccuracy for non-Black participants. It is this race-free CKD-EPI_{cr} equation derived from Inker et al. (2021) that the Task Force recommends for use in clinical settings in the US.

Last, several studies report that the race adjustment in the CKD-EPI_{cr} equation did not improve the accuracy of GFR estimations in Black African populations (e.g., Bukabau et al. 2019). This finding is relevant if, in the US, the racial category 'Black' refers to any persons "having origins in any of the black racial groups of Africa" (Office of Management and Budget 1997, 58789)—in that case, one would expect to see the race adjustment improve the accuracy of GFR estimation *across* the Black racial group (Eneanya, Boulware, and Tsai 2022). Based on this empirical evidence then, one could plausibly affirm that *excluding* a race adjustment in the CKD-EPI_{cr} equation leads to better predictions of GFR, especially for Black patients. Despite the conflicting evidence, clinicians must make a judgment about accepting or rejecting *H*, because as it stands, the CKD-EPI_{cr} equation is the most cost-effective and feasible means of estimating kidney function in the US clinical setting (Delgado et al. 2021).

35.5.2 Inductive risk

I am now able to translate the debate about the race-adjusted CKD-EPI_{cr} equation into the language of inductive risk. My central claim is that eliminativists and conservationists clearly express different values, and these values plausibly play a role in their different conclusions. Although not my focus in this section, it is possible that they may also be differing in how they assess the strength of evidence in favor of accepting *H*. Recall that conservationists argue that we ought to retain the use of race in eGFR_{cr} equations. Conservationists are concerned with more carefully avoiding a false negative about hypothesis *H*—that is, where clinicians wrongly reject that using a race adjustment in the CKD-EPI_{cr} equation generates more accurate estimations of GFR for patients, when it does. Specifically, because they seem to weigh the consequences of wrongly accepting *H* (i.e., accepting a false negative) as particularly worse, conservationists deem the methodological choice to include a race adjustment in the equation as more acceptable (Douglas 2000, 565). To see that this is the case, consider this statement:

We argue to the contrary that simply dropping the race correction factor from eGFR estimates will cause foreseeable harm to Black patients by creating a situation where they regularly receive inappropriate medical treatment based on inaccurate and racially biased estimates of GFR.
(Velasco and Snodgrass 2021, 827)

Conservationists cite Levey, Stevens et al.'s (2009; see also Levey, Tighiouart et al. 2020) studies that the race adjustment in the CKD-EPI_{cr} equation generates more accurate GFR estimations for patients (see Manski 2022; Manski, Mullahy, and Ventakataramani 2023; Velasco and Snodgrass 2021; Williams 2023). For them, erroneously rejecting hypothesis *H* will lead to the regulatory consequence that race is excluded from the CKD-EPI_{cr} equation—just as the NKF-ASN Task Force now recommends for all medical institutions in the US (Delgado et al. 2021).

However, for conservationists, wrongly rejecting *H* is undesirable as it may result in inaccurate eGFR assessment that would harm all patients by leading to inappropriate clinical care (Manski 2022; Manski, Mullahy, and Ventakaramani 2023; Velasco and Snodgrass 2021; Williams 2023). For example, Diao et al. (2021) investigate the possible effects on US clinical care of adopting Inker et al.'s (2021) race-free CKD-EPI_{cr} equation. Williams (2023)—a conservationist—uses Diao et al.'s (2021) projections to claim that the race-free equation will likely underestimate the eGFR of Black patients, and thereby erroneously promote new or worse diagnoses of CKD in approximately 1-million Black patients. Williams (2023) insists that excluding race will likely overestimate eGFR for non-Black patients and thereby lead to overdosing of drugs like metformin in 1.47 million of these patients. Conservationists value these effects as especially worse because they could foreseeably: Increase the CKD public health burden; negatively affect the clinical care of millions of CKD patients; and exaggerate health disparities between Black and non-Black patients, amongst others (Manski 2022; Velasco and Snodgrass 2021, 828; Williams 2023).

To avoid a false negative then, if there is some evidence that the race adjustment improves the accuracy of GFR estimations, conservationists seem to be willing to go with it. Notably, conservationists hold the inclusion of race in the CKD-EPI_{cr} equation to the same evidential standard as any other predictor in a clinical predictive algorithm.

The clinical prediction model literature teaches developers to include in their predictive algorithms only those variables that are strong predictors of a health outcome. These are variables that are readily available in the dataset, easy to measure, and statistically associated with the outcome (Steyerberg 2019, 48). In their study, Levey, Stevens et al. (2009) found that race was a strong predictor of eGFR, and they included race in the CKD-EPI_{cr} equation because they observed higher serum creatinine levels among their Black study participants compared to non-Black participants. Levey et al. also found that including race in the equation produced eGFR estimates that more closely matched study participants' mGFR values than a race-free equation. Conservationists cite these as sufficient reason for clinicians to include a race adjustment in the CKD-EPI_{cr} equation.

Eliminativists question Levey, Stevens et al.'s findings (2009) (see also Levey, Tighiouart et al. (2020)). These critics appear to worry about more carefully avoiding a false positive about *H*—that is, where clinicians wrongly accept that using a race adjustment in the CKD-EPI_{cr} equation generates accurate estimations of GFR, when it does not. Specifically, eliminativists want to see much more evidence to justify using a race adjustment in the CKD-EPI_{cr} equation because they weigh the social impacts of significantly overestimating the eGFR for Black patients as much worse. For them, if in doubt, it is more apt to exclude the race adjustment from the CKD-EPI_{cr} equation.

First, eliminativists worry that clinicians may err in including a race adjustment for some of the following reasons (see Williams (2023) for a response). Recall that Levey, Stevens et al. (2009) speculate that the observed racial differences in serum creatinine levels are caused by Black people having greater muscle mass than non-Black people—their so-called muscle mass hypothesis (Roberts 2021). Eliminativists take this hypothesis to be the *crux* of the conservationist argument: The physiological explanation for the race adjustment in the CKD-EPI_{cr} equation.

However, eliminativists contend that the race correction in the CKD-EPI_{cr} equation is not based on “robust evidence” because the muscle mass hypothesis is scientifically flawed (Braun et al. 2021; Eneanya, Boulware, and Tsai 2022; Roberts 2021). This is another inductive risk issue in the debate. Levey et al. support the muscle mass hypothesis by citing three studies (see Cohn et al. 1977; Harsha, Frerichs and Berenson 1978; Worrall et al. 1990). But eliminativists think we need more evidence to support the hypothesis. For these critics, not only are the three studies based on an unrepresentative sample of 176 Black participants in total, but none of them directly prove the muscle mass hypothesis (Franks and Scott 2021). In fact, other studies like Hsu, Johansen et al.

(2008) find that Black patients do have higher serum creatinine levels than similar White patients, but not because of muscle mass.

Eliminativists also question the need for a race adjustment because they claim that Levey, Stevens et al.'s (2009) original CKD-EPI study was based on an unrepresentative population of Black US adults (Eneanya, Boulware, and Tsai 2022; Franks and Scott 2021; Mohottige et al. 2022). Nearly 70% of the Black participants in that study already had extensive kidney disease and were especially impoverished. According to Eneanya and colleagues (2022, 88), for example, muscle mass notwithstanding, one should *expect* to see elevated serum creatinine levels in this Black population, compared to the non-Black study population which was composed of participants with a greater variety of CKD and normal kidney function.

Moreover, eliminativists insist that including a race adjustment in the CKD-EPI_{cr} equation risks erroneously using race as a “biological variable” (Braun et al. 2021; Delgado et al. 2021; Eneanya, Boulware, and Tsai 2022; Vyas, Eisenstein, and Jones 2020). The idea that race is biologically real is that there is a way to define race using a biological property like genetic differences or common ancestry, that is consistent with what contemporary scientific evidence says about human races. For eliminativists, if we have a biological conceptualization of “race” in clinical algorithms, allowing that observed racial differences in health risks are natural, then this ignores social mechanisms—including the effects of structural racism—that are more likely to explain the racial differences. For them, one race adjustment for all Black people obscures the social and genetic heterogeneity amongst this group (Braun et al. 2021).

Second, eliminativists *explicitly* trade-off the consequences of a false negative and false positive about hypothesis *H* and deem the latter as more undesirable. For example, Eneanya, Boulware, and Tsai (2022, 90) acknowledge that erroneously removing the race adjustment in the CKD-EPI_{cr} equation risks overdiagnosing CKD in Black patients and reversing CKD diagnoses and treatment for many non-Black patients. But they believe the risk of overestimating eGFR in Black patients by using the race adjustment is worse:

By contrast, the potential advantages of removing the race coefficient include earlier awareness of CKD [for Black patients] and more timely referral to general and transplant nephrology care for prevention or treatment.

(Eneanya, Boulware, and Tsai 2022, 90)

Eliminativists highlight the consequences of wrongly accepting *H* for Black patients *because* this population has a higher incidence of CKD and end-stage renal disease in the US Black patients with CKD are also less likely to be referred to kidney specialists and waitlisted for kidney transplants (Purnell, Luo, and Cooper 2018). So, for eliminativists, if there's a chance that the race adjustment overestimates Black patients eGFR and thereby *further* delays their CKD treatment, then this too greatly risks exacerbating systemic health inequities in the US and unjustly disadvantaging an already marginalized patient population (see Delgado et al. 2021; Eneanya, Boulware, and Tsai 2022, 84).

Importantly, eliminativists are also concerned that wrongly using a race adjustment could perpetuate pernicious values in medical research, education, and clinical practice such as race essentialism, stereotyping, and bias, thereby undermining health equity. For them, race correction is an exemplar of race-based medicine. Cerdeña and colleagues (2020) describe race-based medicine as a health system that ignores the effects of structural racism on health; reinforces harmful stereotypes and biologized notions of race by portraying certain racial and ethnic groups as “high risk” or “inherently diseased”; and increases health inequities by erroneously narrowing diagnoses, restricting treatment options,

and hindering individualized care for minority patients. By contrast, conservationists do not seem as worried about these possible effects of using a race adjustment.

Weighing these non-epistemic consequences of error as especially serious then, [Vyas, Eisenstein, and Jones \(2020\)](#) offer three questions clinicians should ask before developing and deploying race-adjusted clinical algorithms:

Is the need for race correction based on robust evidence and statistical analyses (e.g., with consideration of internal and external validity, potential confounders, and bias)? Is there a plausible causal mechanism for the racial difference that justifies the race correction? And would implementing this race correction relieve or exacerbate health inequities? (880)

These questions just are an inductive risk calculation. For [Vyas, Eisenstein, and Jones \(2020\)](#), even if a race correction seems somewhat helpful, there needs to be much more evidence (including a causal mechanism) to justify using it. Additionally, in their assessment of the evidence on 26 approaches to GFR estimation, and their recommendation to use the race-free 2021 CKD-EPI_{cr} equation, the NKF-ASN Task Force was also engaged in an inductive risk calculation. The Task Force is explicit that the accuracy of the eGFR equations formed only part of their deliberations. Whether the winning equation was also derived in a demographically diverse population, could be easily implemented, is patient centered, avoids foreseeable adverse clinical consequences, and did not disproportionately impact any one group, was also part of their evidential deliberation ([Delgado et al. 2021](#)). The Task Force states that the winning race-free CKD-EPI_{cr} equation best matched the “scientific evidence and values held by members” ([Delgado et al. 2021](#), 2295). The Task Force’s decision is fueled in part by their weighting of the drawbacks of wrongly using race in GFR estimation as worse.

35.6 Values and race correction

I have just shown that as with classic cases of inductive risk, non-epistemic values seem to be inseparable from the methodological choice to include or exclude the race adjustment in the CKD-EPI_{cr} equation. Using [Ward’s \(2021\)](#) taxonomy, it appears that scientists are consciously motivated by values in making the methodological choice (values act as causal effectors), and the choice itself profoundly impacts things we value (it promotes specific values over others). My analysis on inductive risk in the race correction debate has several implications for philosophical discussions about values in science.

First, [Douglas \(2009\)](#) emphasizes that gathering a greater *quantity* of evidence lessens inductive risk and thus makes it easier to manage. But just as in [Bluhm’s \(2017\)](#) analysis of clinical research, it is not clear that the *amount* of evidence is the sole or primary issue involved in addressing inductive risk in the race correction debate. For example, when deliberating whether to accept hypothesis *H*, eliminativists are concerned with not just how much evidence there is (e.g., evidence of the accuracy of the CKD-EPI_{cr} equation) but also with the *quality* of the evidence ([Bluhm 2017](#), 208). This issue of “quality” includes which studies even count as relevant evidence. For example, we know that to justify including a race adjustment, [Vyas, Eisenstein, and Jones \(2020\)](#) want to see evidence of a causal mechanism. Moreover, eliminativists claim that the studies [Levey, Stevens et al. \(2009\)](#) cite do not count as decent or relevant evidence of the muscle mass hypothesis. In its deliberations, the NKF-ASN Task Force also considered testimony from patients about the “patient centeredness” of various approaches to GFR estimation ([Delgado et al. 2021](#)). This qualitative

dimension means a greater quantity of evidence about H would not necessarily solve the race correction debate (Bluhm 2017).

Second, about the Task Force's decision to recommend race-free GFR estimation, Williams (2023) charges that "[t]he exclusion of race was a pre-condition to the choice of formulas and not the consequence of rigorous statistical evaluation." We know that the Task Force considered several kinds of evidence, including "rigorous statistical evaluation." Yet there is a point to be made that no matter the quantity or quality of evidence gathered about H , some eliminativists are weary about any use of race in medicine *at all* because they think it will promote pernicious values like race essentialism. That is, it looks like for some eliminativists, there is no scenario in which race correction would be appropriate because "race" is too problematic to use in science (e.g., Cerdeña, Plaisime, and Tsai 2020; Roberts 2021). For these kinds of eliminativists then, one could argue that ethical values *are* playing a large role in methodological choice by completely *ruling out* race correction as acceptable. This type of eliminativist move appears to be what Douglas calls a "direct role" for values—where values serve as reason to accept a claim (which she regards as unacceptable)—as opposed to a typical "indirect role" for values which she accepts, where values are used only to set evidential standards (Douglas 2009, 96; see also Elliott (2011)). Future research could investigate whether ethical values *are* playing a direct role in methodological choice for these eliminativists and if, whatever the role is, it is problematic or *warranted*.

Note

- 1 To be sure, although Rudner (1953) makes use of the concept of inductive risk, the term "inductive risk" was first coined by Hempel (1954).

References

- Biddle, Justin B. 2016. "Inductive Risk, Epistemic Risk, and Overdiagnosis of Disease." *Perspectives on Science* 24 (2): 192–205. https://doi.org/10.1162/POSC_a_00200.
- Bluhm, Robyn. 2017. "Inductive Risk and the Role of Values in Clinical Trials." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press.
- Braun, Lundy, Anna Wentz, Reuben Baker, Ellen Richardson, and Jennifer Tsai. 2021. "Racialized Algorithms for Kidney Function: Erasing Social Experience." *Social Science & Medicine* 269: 113548. <https://doi.org/10.1016/j.socscimed.2020.113548>.
- Bukabau, Justine B., Eric Yayo, Appolinaire Gnionsahe, et al. (2019). "Performance of Creatinine- or Cystatin C-based Equations to Estimate Glomerular Filtration Rate in Sub-Saharan African Populations." *Kidney International*, 95: 1181–89. <https://doi.org/10.1016/j.kint.2018.11.045>.
- Cerdeña, Jessica P., Marie V. Plaisime, and Jennifer Tsai. 2020. "From Race-Based to Race-Conscious Medicine: How Anti-Racist Uprisings Call Us to Act." *The Lancet* 396 (10257): 1125–8.
- Cohn, Stanton H., Concepcion Abesamis, Italo Zanzi, John F. Aloia, Seiichi Yasumura, and Kenneth J. Ellis. 1977. "Body Elemental Composition: Comparison between Black and White Adults." *American Journal of Physiology: Endocrinology and Metabolism* 232 (4): E419–22. <https://doi.org/10.1152/ajpendo.1977.232.4.E419>.
- Delgado, Cynthia, Mukta Baweja, Deidra C. Crews, et al. 2021. "A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease." *Journal of the American Society of Nephrology* 32 (12): 2994–3015.
- Diao, James A., Gloria J. Wu, Herman A. Taylor, et al. 2021. "Clinical Implications of Removing Race from Estimates of Kidney Function." *Journal of the American Medical Association* 325 (2): 184–6. <https://doi.org/10.1001/jama.2020.22124>.
- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.

- Elliott, Kevin C. 2011. "Direct and Indirect Roles for Values in Science." *Philosophy of Science* 78 (2): 303–24. <https://doi.org/10.1086/659222>.
- Eneanya, Nwamaka D., L. Ebony Boulware, Jennifer Tsai, et al. 2022. "Health Inequities and the Inappropriate Use of Race in Nephrology." *Nature Reviews Nephrology* 18: 84–94. <https://doi.org/10.1038/s41581-021-00501-8>.
- Franks, Caroline E., and Mitchell G. Scott. 2021. "On the Basis of Race: The Utility of a Race Factor in Estimating Glomerular Filtration." *The Journal of Applied Laboratory Medicine* 6 (1): 155–66. <https://doi.org/10.1093/jalm/jfaa128>.
- Harsha, David W., Ralph R. Frerichs, and Gerald S. Berenson. 1978. "Densitometry and Anthropometry of Black and White Children." *Human Biology* 50 (3): 261–80. <https://www.jstor.org/stable/41463071>.
- Hempel, Carl G. 1954. "A Logical Appraisal of Operationism." *The Scientific Monthly* 79 (4): 215–20. <https://www.jstor.org/stable/21072>.
- Hsu, Chi-yuan, Wei Yang, Rishi V. Parikh et al. 2021. "Race, Genetic Ancestry, and Estimating Kidney Function in CKD." *New England Journal of Medicine* 385: 1750–60. <https://doi.org/10.1056/NEJMoa2103753>.
- Hsu, Joy, Kirsten L. Johansen, Chi-yuan Hsu, George A. Kaysan, and Glenn M. Chertow. 2008. "Higher Serum Creatinine Concentrations in Black Patients with Chronic Kidney Disease: Beyond Nutritional Status and Body Composition." *Clinical Journal of the American Society of Nephrology* 3 (4): 992–7. <https://doi.org/10.2215/CJN.00090108>.
- Inker, Lesley A., Nwamaka D. Eneanya, Josef Coresh, et al. 2021. "New Creatinine- and Cystatin C-Based Equations to Estimate GFR without Race." *New England Journal of Medicine* 385 (19): 1737–49. <https://doi.org/10.1056/NEJMoa2102953>.
- Levey, Andrew S., Hocine Tighiouart, Silvia M. Titan, and Lesley A. Inker. 2020. "Estimation of Glomerular Filtration Rate with vs without Including Patient Race." *Journal of the American Medical Association Internal Medicine* 180 (5): 793–5. <https://doi.org/10.1001/jamainternmed.2020.0045>.
- Levey, Andrew S., Lesley A. Stevens, Christopher H. Schmid, et al. 2009. "A New Equation to Estimate Glomerular Filtration Rate." *Annals of Internal Medicine* 150 (9): 604–12. <https://doi.org/10.7326/0003-4819-150-9-200905050-00006>.
- Levey, Andrew S., Silvia M. Titan, Neil R. Powe, Josef Coresh, and Lesley A. Inker. 2020. "Kidney Disease, Race, and GFR Estimation." *Clinical Journal of the American Society of Nephrology* 15: 1203–12.
- Lusk, Greg, and Kevin C. Elliott. 2022. "Non-Epistemic Values and Scientific Assessment: An Adequacy-for-Purpose View." *European Journal for Philosophy of Science* 12 (2): 35. <https://doi.org/10.1007/s13194-022-00458-w>.
- Manski, Charles F. 2022. "Patient-Centered Appraisal of Race-Free Clinical Risk Assessment." *Health Economics* 31 (10): 2109–14. <https://doi.org/10.1002/hec.4569>.
- Manski, Charles F., John Mullahy, and Atheendar S. Ventakaramani. 2023. "Using Measures of Race to Make Clinical Predictions: Decision Making, Patient Health, and Fairness." *PNAS* 120 (35): e2303370120. <https://doi.org/10.1073/pnas.2303370120>.
- Marzinko, Mark A., Dina N. Greene, Patrick M. Bossuyt, et al. 2022. "Limited Evidence for Use of a Black Race Modifier in eGFR Calculations: A Systematic Review." *Clinical Chemistry* 68 (4): 521–33. <https://doi.org/10.1093/clinchem/hvab279>.
- Mohottige, Dinushika, L., Ebony Boulware, Chandra Ford, Camara Jones, and Keith C. Norris. 2022. "Use of Race in Kidney Research and Medicine: Concepts, Principles, and Practice." *Clinical Journal of the American Society of Nephrology* 17 (2): 314–22. <https://doi.org/10.2215/CJN.04890421>.
- Office of Management and Budget. 1997. "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity." *Federal Register: The Daily Journal of the United States Government* 62 (210): 58782–90.
- Parker, Wendy S., and Eric Winsberg. 2018. "Values and Evidence: How Models Make a Difference." *European Journal for Philosophy of Science* 8 (1): 125–42. <https://doi.org/10.1007/s13194-017-0180-6>.
- Purnell, Tanjala S., Xun Luo, Lisa A. Cooper, et al. 2018. "Association of Race and Ethnicity with Live Donor Kidney Transplantation in the United States from 1995 to 2014." *JAMA* 319 (1): 49–61. <https://doi.org/10.1001/jama.2017.19152>.
- Roberts, Dorothy E. 2021. "Abolish Race Correction." *Lancet* 397 (10268): 17–8.
- Rudner, Richard. 1953. "The Scientist Qua Scientist Makes Value Judgements." *Philosophy of Science* 20: 1–6. <https://doi.org/10.1086/287231>.
- Steyerberg, Ewout W. 2019. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. 2nd ed. Springer.
- Velasco, Joel D., and Brad Snodgrass. 2021. "The Use of Race in eGFR: Why Racial Justice Requires Accuracy." *The American Journal of Medicine* 134 (7): 827–8.

- Vyas, Darshali, Leo G. Eisenstein, and David S. Jones. 2020. "Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms." *New England Journal of Medicine* 383 (9): 874–82. <https://doi.org/10.1056/NEJMms2004740>.
- Ward, Zina B. 2021. "On Value-laden Science." *Studies in History and Philosophy of Science* 82 (2): 54–62. <https://doi.org/10.1016/j.shpsa.2020.09.006>.
- Williams, Paul. 2023. "Retaining Race in Chronic Kidney Disease Diagnosis and Treatment." *Cureus* 15 (9): e45054. <https://doi.org/10.7759/cureus.45054>.
- Worrall, John Gatlan, Virach Phongsathorn, Richard J. L. Hooper, and Elisabeth W. Paice. 1990. "Racial Variation in Serum Creatine Kinase Unrelated to Lean Body Mass." *British Journal of Rheumatology* 29 (5): 371–3. <https://doi.org/10.1093/rheumatology/29.5.371>.
- Zelnick, Leila R., Nicolae Leca, Bessie Young, and Nisha Bansal. 2021. "Association of the Estimated Glomerular Filtration Rate with vs without a Coefficient for Race with Time to Eligibility for Kidney Transplant." *JAMA Network Open* 4 (1): e2034004. <https://doi.org/10.1001/jamanetworkopen.2020.34004>.

VALUE OVERLAPS IN TRANSDISCIPLINARY AND INTERCULTURAL COLLABORATION

Charbel N. El-Hani and David Ludwig

36.1 Partial overlaps and the value-ladenness of science

In previous works, we developed a “partial overlaps” framework that emphasizes, on the one hand, overlaps between epistemologies, ontologies, and values included in different knowledge systems, and, on the other, differences that make those overlaps always partial. Overlaps can act as springboards for reaching common ground and coordinated joint action, while their partialities require serious engagement with disagreements and tensions (Ludwig 2016a; Ludwig and El-Hani 2020, 2025). The recognition of partial overlaps is particularly relevant in transdisciplinary contexts that involve diverse academic and non-academic actors and in intercultural contexts that involve actors of diverse cultural backgrounds. While the assumption of full incommensurability can lead to the exclusion and exoticization of actors, an emphasis on overlaps allows for the exploration of similarities between knowledge systems that can ground mutual understanding and collaboration. While the assumption of full commensurability can obscure differences and assimilate marginal standpoints into dominant frameworks, the emphasis on partialities allows for recognition of diverging perspectives and mechanisms of exclusion.

Partial overlaps make it possible, thus, that holders of distinct knowledge systems search for possibilities of engaging with each other’s perspectives while avoiding denying differences in their conceptions and assumptions. A requisite for fruitfully relating knowledge systems, however, is to ascertain how they can be identified. Knowledge systems are built by communities as networks of agents interconnected by social relationships through which they dynamically combine doing, learning, and knowing (van Kerkhoff and Szlezák 2016), and eventually establish practices that mediate knowledge production, transfer, and use (Cornell et al. 2013). These systems show a determinate degree of internal and practical coherence, derived from the way they have been developed by generations of agents dealing with situations embedded in the socio-environmental circumstances in which they carry out their cognitive and practical actions. They show, thus, pragmatic value in relation to those circumstances and the fruitfulness of putting them into dialogue derives from the fact that human communities face a diversity of challenges, which entailed the development of different conceptions and assumptions showing pragmatic value.

The aim of this chapter is to introduce the partial overlaps framework with particular attention to the roles of values in transdisciplinary and intercultural collaboration. Philosophy of science widely embraces the value-ladenness of science (Brown 2020; Douglas 2009; Elliott 2017) while

simultaneously emphasizing the importance of social diversity in research (Harding 2019; Ludwig and Ruphy 2024; Rolin 2019). However, value-ladenness and social diversity together create complex landscapes of heterogeneous and sometimes conflicting values that need to be navigated in scientific practice (van de Poel 2015). As these challenges are particularly striking in the context of transdisciplinary and intercultural collaboration (Milberg Muñiz, Ludwig, and El-Hani 2025), this chapter discusses how the partial overlaps framework allows for dialogues and mutual understanding across knowledge systems that are marked by value diversity.

The next section situates the partial overlaps framework in a wider history of epistemic paternalism, diversity, and decolonization in transdisciplinary practice (Ludwig and El-Hani 2025). It is argued that the paternalistic legacy of international development is grounded in the legacy of the value-free ideal while more recent debates about epistemic diversity and epistemic decolonization highlight the entanglement of values and knowledge production. The third section focuses on partial overlaps between values articulated in different knowledge systems, especially focusing on Indigenous and local practices and knowledge. Finally, the fourth section considers the possibilities of learning from partial overlaps, particularly in the entanglement of values and ontologies.

36.2 Paternalism, diversity, and decolonization in the politics of knowledge

International development emerged after World War II as a paternalist project that assumed that science and technology of the “developed world” could be simply exported into the “underdeveloped world,” without recognizing the epistemic agency of the latter, and purportedly generating economic growth and societal progress. These projects were part of an effort to metamorphosize colonialism into developmentalism, as shown by Harry S. Truman’s 1949 inaugural speech as US president. Arguing that the “old imperialism—exploitation for foreign profit—has no place in our plans,” Truman promised a new phase of global progress that would make “the benefits of our scientific advances and industrial progress available for the improvement and growth of underdeveloped areas” (see Binns 2014). The rhetoric of development aimed at helping an “underdeveloped world” through paternalistic science and technology remained very much in the epistemic tradition of colonial science, as pursuing a “civilizing mission,” highlighting the expertise of the colonizers while treating the colonized as passive beneficiaries with no epistemic agency and very limited decision-making power.

The idea of a value-free science was part of the developmentalist perspective, which demanded protecting research from political and societal influence, as articulated in Merton’s (1979/1942) “norms of science.” A supposedly depoliticized “front-end” of science (through disinterestedness, universality, objectivity, neutrality, and value-freedom) was combined with a political “back-end” of societal benefits (through development, growth, modernization, progress) (Ludwig and El-Hani 2025). However, once one takes into account the role of values in scientific and technological research, a critical perspective on epistemic paternalism naturally follows. As science cannot be exported in a value-free manner, paternalistic narratives of international development often amount to impositions of dominant values and future visions onto allegedly underdeveloped countries. In converse, critical reflections on the relations between science and values become central to alternative approaches to development, for instance, by recognizing the role of epistemic diversity in knowledge production, as enacted in transdisciplinary research, or engaging with decolonial considerations on the power dimension of relations between science, technology, and society.

While the field of international development often remains deeply paternalistic (Boogaard 2021), growing debates about both epistemic diversity and epistemic decolonization reflect a “new politics of knowledge” (Ludwig and El-Hani 2025), which commonly appeals to “transdisciplinarity” as an umbrella term. The notion of transdisciplinarity often remains ambiguous and vague, while being linked to a plethora of buzzwords such as “citizen science,” “co-production,” “community-based

research,” “community-driven development,” “intercultural dialogue,” “multi-stakeholder approaches,” “participatory research,” “responsible research and innovation,” “open science,” among many others (e.g., [Caniglia and Russo 2024](#); [Leonelli 2023](#); [Ludwig and Boogaard 2021](#); [van Maurik Matuk et al. 2023](#)).

These perspectives often assume the aim of “integrating” diverse forms of academic and non-academic knowledge. Epistemically, the promise is that transdisciplinary knowledge integration can align epistemic and political concerns of knowledge production, addressing complex challenges that require diverse forms of expertise of both academic and non-academic actors. Transdisciplinarity is not limited, however, to epistemic promises, as it is typically intertwined with a political promise of overcoming paternalistic structuring of the interface of science and society. Integrating diverse forms of knowledge also means integrating diverse values and thereby allowing for more inclusive forms of doing science. Especially in the Global South, transdisciplinarity is called to conjoin epistemic and political ambitions through inclusive knowledge production that takes the knowledge and livelihood concerns of disenfranchised communities into account.

There is always a risk, however, of a tame appeal to diversity that recognizes other forms of knowledge primarily as data for academic consumption ([Ludwig and Boogaard 2021](#)), resulting in “knowledge extraction” or “knowledge mining” ([Alcoff 2022](#); [Kimmerer 2012](#)). In this manner, acknowledging that, say, Indigenous or local communities have useful knowledge that can be accessed through transdisciplinary research may not translate into political support for their struggles for self-determination. The political promise that transdisciplinarity might lead to epistemically robust and socially just research often clash with academic structures that tend to assimilate diversity in superficial ways for managerial and legitimizing purposes ([De La Rosa et al. 2024](#); [Milberg Muñiz, Ludwig, and El-Hani 2025](#)).

The literature on science and values provides a helpful lens for understanding why the shift from epistemic paternalism to epistemic diversity is often not sufficient to avoid extractivist dynamics. If Indigenous and local knowledge are integrated into academic knowledge, the former often becomes decontextualized and put into service of frameworks that are shaped by the interests and values of academic researchers. For example, a conservation biologist may recognize the usefulness of Indigenous and local knowledge for their own goal of conserving a critically endangered species while not substantially engaging with the goals and values of Indigenous and local communities. Knowledge integration can therefore obscure value diversity while prioritizing the goals of dominant actors. Understanding knowledge systems as value-laden creates an alternative entry point: Rather than asking how Indigenous and local knowledge can be useful for academic goals, it points to the question of how diverse knowledge systems support equally diverse goals that are grounded in different value systems.

Decolonial scholarship (e.g., [Chilisa 2019](#); [Cusicanqui 2012](#); [Smith 1999](#)) responds to these limitations of knowledge integration by focusing on differences and inequalities among actors. Rather than presenting a picture of harmonious knowledge integration, decolonial scholars stress the need to take others seriously even when their ontologies, epistemologies, values are at odds with dominant ones. This shift from epistemic diversity to epistemic decolonization is often driven by critiques of capitalist, colonial, and patriarchal values that underlie dominant forms of academic knowledge production ([Escobar 2020](#)). Decolonial scholarship emphasizes that academic knowledge production commonly masks its oppressive values through the ideal of value-freedom—for example, when valuing nature primarily as a resource for commodity production while silencing alternative ways of valuing nature in non-instrumental and non-commodified ways ([Büscher and Fletcher 2020](#); [Moore 2015](#)). Indigenous and local communities are therefore not merely portrayed as holding oppressed knowledge but also oppressed values, for example, when relating to animals and plants as kin that demand respect and reciprocity rather than as resources that can be exploited

for commercial profit, even if sustainably, or when focusing on their own livelihood concerns rather than external agendas of economic growth. Celebrating epistemic diversity is not sufficient and can reinforce oppressive dynamics when Indigenous and local knowledge are integrated into frameworks that are shaped by the values of external actors rather than Indigenous and local communities themselves.

The shift from epistemic diversity to epistemic decolonization highlights the shortcoming of a simple integrationism while centering on differences between knowledge systems and values. An exclusive focus on difference, however, can also backfire in creating the illusion of full incommensurability that undermines any prospects of constructive collaboration and dialogue in the face of inequality and oppression. It may leave little room for positive visions of intercultural exchange, knowledge co-production, and value plurality. There is a productive tension, then, between a criticism of epistemic paternalism from the perspective of an appeal to diversity, which emphasizes collaboration between different knowledge systems, and a decolonial critique of both paternalism and transdisciplinarity that stresses tensions and differences that should not be disguised by a rhetorical emphasis on similarities and synergies. It is from this productive tension that a transformative transdisciplinary approach is proposed, which moves beyond an epistemic paternalism that exclusively relies on academic knowledge by combining an appreciation of epistemic diversity and a critique of epistemic inequality (Ludwig and El-Hani 2025).

It is important to develop epistemological and methodological frameworks that take into account both the possibilities and limitations of intercultural dialogue, knowledge co-production, and value plurality. That's why the partial overlaps framework, which proposes an analysis of the relations between knowledge systems in terms of convergences (*overlaps*) and divergences (*partialities*) in their ontologies, epistemologies, and values, is a central element in transformative transdisciplinarity (Ludwig and El-Hani 2020, 2025). On the one hand, ontological, epistemological, and value overlaps can show or promote common grounds or standpoints that encourage collaboration and mutual learning, opening a space for shared experience, intercultural translation, knowledge co-production, and so forth. On the other hand, significant differences in ontologies, epistemologies, and values can foster reflection among heterogeneous social actors, including researchers and Indigenous/local communities, about their normative and political positionings in connection to the diversity of knowledge systems and their equally diverse relations. This reflection can provoke, then, an engagement with the construction of a political philosophy of intercultural relations, which may contribute to the equity and quality of transformative transdisciplinary processes in situated settings. Moreover, a thoughtful analysis of partial overlaps between knowledge systems can inspire ways of learning from difference, even if radical, through intercultural translation (El-Hani and Ludwig 2025). The partial overlaps framework creates conditions for shifting from a recognition of difference to the search for convergences that can offer a point of departure for dialogue between knowledge systems, for attempts at intercultural translation, and for efforts to deal with distinct ontological, epistemological, and value commitments without losing from sight the integrity, autonomy, and self-determination of different cultures and knowledge systems.

36.3 Partial overlaps in value commitments

When inquiring into partial overlaps, it is important to consider that ontological, epistemological, and value commitments co-determine one another. These commitments are entangled in knowledge systems and so are their partial overlaps. Thus, an inquiry into partial overlaps can focus on any of these commitments and will end up showing an entanglement of ontology, epistemology, and values. To begin by focusing on ontological, epistemological or value partial overlaps is mostly a methodological decision on how to approach the issue of convergences and divergences between knowledge

systems. Thus, while this section will focus on values, it will not be possible to discuss them separately from ontology and epistemology.

As a case in point, let us consider how Indigenous and local practices are often valued as inspirations and models for sustainability. Since the *Brundtland Report*, which first introduced the idea of sustainable development, it has been argued that tribal and Indigenous “lifestyles can offer modern societies many lessons in the management of resources in complex forest, mountain and dryland ecosystems” (WCED 1987, 12). The idea of sustainability is, thus, relatively recent, even though some sustainability-related ideas can be traced back to early modern (Caradonna 2022) or even ancient (Gomis et al. 2011) philosophy. Meanwhile, many of the Indigenous and local practices described as “sustainable” by academics are centuries or even millennia old. Therefore, to characterize them as sustainable is an act of intercultural translation that allows proposing overlaps with a guiding principle that has become rather dominant at the intersection of normative and empirical concerns about socio-ecological dynamics in the last half century.

Sustainability discourses commonly appeal to an overlap of values. Actors like Indigenous communities and academic researchers differ in many ways but find common ground in valuing sustainability in the sense of non-destructive relations between people and their environments. This claim of a value overlap can motivate optimistic visions of co-creation, co-production, co-management, etc. involving academic scientists, policymakers, Indigenous groups, and others, under the motivation of converging reasoning about moral responsibilities toward environments. As such, appeals to sustainability encourage academic research to move closer to Indigenous and local knowledge. The recognition of such overlaps can even support meaningful co-creation that recognizes holders of Indigenous and local knowledge as experts who can teach scientists how to weave moral responsibility and environmental stewardship into biological and ecological research. But this expectation ultimately shows how that recognition comes with challenges related to differences between knowledge systems, i.e., to partialities of overlap. Consider, for instance, that Indigenous and local communities do not assume a clear dichotomy between “facts” and “values,” as they do not assume a separation between nature and culture (e.g., Descola 2005), and treat knowledge about environments as a fundamentally moral issue, in contrast to the commitment to value-free perspectives still found in scientific communities to a significant extent.

There is certainly some truth to the optimistic vision of a shared concern about sustainability, which can play an important role in creating awareness for Indigenous and local values in conservation practices and environmental policy. At the same time, stories about overlap or complementarity in sustainability can be misleading and outright harmful if they contribute to a neglect of tensions between heterogeneous value systems. First, environmental practices of Indigenous and local communities are highly diverse and their blanket characterizations as sustainable can mislead about their complexity through a simplistic and romanticized picture of sustainability. Second, characterizations of Indigenous and local knowledge and practices as sustainable also require critical reflectivity about how different value systems can lead to different understandings and operationalizations of “sustainability.”

Four primary dimensions of sustainable development can be identified in the *Brundtland Report* (Holden, Linnerud, and Banister 2014): “safeguarding long-term ecological sustainability, satisfying basic human needs, and promoting intragenerational and intergenerational equity.” While these dimensions suggest overlaps with value systems of Indigenous and local communities, salient differences can also be discerned. Indigenous knowledge, for instance, often assumes a complex web of mutual responsibilities between human and non-human actors (Lewis-Jones 2016; Rose 2002) that is not adequately reflected in satisfying basic human needs or human-focused equity concerns. As non-human actors are incorporated into the value outlooks of Indigenous communities with both responsibilities and rights, their normative reasoning about environmental issues will often depart from

contemporary discourses about sustainable development. Moreover, sustainability discourses tend to be geared to sustainable growth that is expected to remain compatible with long-term ecological sustainability and intergenerational equity. Meanwhile, Indigenous people typically are not committed to economic growth as a central requirement for a good living, such that they would then aim for compatibility with socio-ecological concerns (Elkington 2013). Sustainability is a deeply modern notion that emerges from capitalist understandings of nature as a depletable resource that needs to be rationally managed and is valued for its instrumental usefulness in satisfying human needs. For many Indigenous communities, however, this is a very alien way of thinking about and valuing nature. Understanding nature as a depletable resource is not a universal view and there are, accordingly, important differences in the values involved in rational management practices aiming at sustainable use of resources and the relations established by many local and Indigenous communities with nature.

Investigating partial overlaps in value systems provides a helpful starting point for avoiding both overly optimistic visions of knowledge integration that may obscure differences and reproduce hierarchies between scientists and Indigenous/local communities and overly pessimistic views that exclusively focus on incommensurability and may block the way to intercultural translation, dialogue and learning. On the one hand, the recognition of overlaps avoids a simplistic divide between Indigenous and modern/contemporary values. Postulating an incommensurable divide is often factually incorrect, as Indigenous/local communities and academic researchers share substantive concerns such as improving local livelihoods or preserving biocultural heritage. Furthermore, neglecting overlaps in value systems can also promote the marginalization of Indigenous and local knowledge by rejecting any common ground for meaningful collaboration in the negotiation of practice and policy. On the other hand, the partiality of such overlaps needs to be addressed carefully to avoid illusions of co-creation in cases where Indigenous and local knowledge are being simply mined or assimilated into normative agendas of external researchers and conservation managers. Taking transdisciplinary and intercultural collaboration seriously requires navigating a complex web of relations among values, interests, practices that can provide resources for joint action but also deep normative tensions that need to be taken seriously.

36.4 Learning from partial overlaps

Transdisciplinary and intercultural collaboration demands that heterogeneous social actors learn from each other to become capable of coordinating their actions for dealing with current challenges, from environmental conservation to public health. But how can we learn from each other when we have so many different perspectives, values, knowledge, practices, some of them based on diverging commitments? The partial overlaps framework provides an interesting window to think about learning across knowledge systems.

The diversity of ontologies we find in plural knowledge systems entails a plurality of worldviews, because it entails a diversity of categorical schemes to treat singular objects and events found in experience as “one-of-many-kinds,” as expressed in different languages, habits, and practices (Dewey 1938). Ontologies lead to distinct assumptions and hypotheses about the other entities with which we share the world and how we should act in relation to them to fulfill our interests and satisfy our values. As heterogeneous knowledge systems are entangled with distinct worldviews and categorical schemes, they are committed to different and sometimes conflicting assumptions about “how the world works” and how the communities holding them can act within its dynamics to bring about outcomes that are positively valued and avoid negatively valued consequences. These values affect, in turn, how different communities build their knowledge systems through inquiry and, consequently, ontologies, epistemologies, and methodologies are entangled in the agency of communities striving to reach their goals in the natural world.

To discuss ontologies entails, therefore, addressing values and norms. Ontologies are not just means of representing the world but demand a richer pragmatic analysis of what they do (Ludwig and El-Hani 2025; Ludwig et al. 2024). When we move from ontology as representation to ontology as practice, acknowledging ontological pluralism entails a normative positioning regarding the epistemic (e.g., understanding of social-environmental systems) and political (e.g., self-determination of Indigenous communities) importance of ontological diversity. Under this lens representational tools are seeing as fundamentally shaped by explanatory goals and priorities, pragmatic values and interests, embodied experiences, and practical needs of communities as epistemic agents. In interdisciplinary and intercultural collaboration, learning from each other requires, then, dealing with plural ontologies and knowledges as practices shaped by a diversity of socio-historically situated values and interests.

In transdisciplinary and intercultural collaboration, ontological diversity offers a fruitful springboard for mutual learning in cases of both overlaps and partialities. Consider, for instance, cross-cultural overlaps in taxonomic categories, as shown in many ethnobiological studies (Berlin 1992). To understand the prevalence of intercultural overlaps in taxonomies, consider how different communities build taxonomic categories based on attention to different properties (Ludwig 2016a; Ludwig and El-Hani 2020, 2025). Different ontologies emerge from practices that respond to different concerns (e.g., explaining evolutionary relations and securing livelihoods through agriculture), which lead to attention to specific sets of properties in entities and events. These practices are not value-free; rather, they involve values and interests that are irreducible components of categorization practices. Biological organisms resemble and differ from each other along many different dimensions and categories of organisms are therefore shaped by both epistemic values (e.g., the relevance of properties for ecological, ethological, genetic explanations) and non-epistemic values (e.g., the relevance of properties for economic, medicinal, spiritual practices) that may point in different directions. A growing body of philosophical scholarship argues that the idea of a purely objective taxonomy is itself part of the questionable legacy of the value-free ideal and needs to be challenged through pluralist accounts that can handle the diverse roles of values in shaping categories (Ludwig 2016b; Reydon and Ereshefsky 2022).

While categories are shaped by values, communities with different values do not always employ radically different taxonomies. On the contrary, ethnobiological research shows that taxonomic overlaps often persist despite value plurality. In these cases of taxonomic overlaps, intercultural dialogue will often be epistemically productive because it provides a more comprehensive account of the taxonomically overlapping item, allowing a larger number of inferences that can be valuable for reaching goals in different communities' practices (Ludwig 2016; Ludwig and El-Hani 2020).

Let us mention an example from our fieldwork with artisanal fishing communities in northeast Brazil (El-Hani, Poliseli, and Ludwig 2022; Ludwig and El-Hani 2020), related to the cross-cultural taxonomic overlap between the academic crab species *Goniopsis cruentata* and the local species *Aratu*. While this overlap between academic and local taxonomies is mediated by joint recognition of properties such as a distinct morphology and behavior that distinguishes the species from other crabs, local fishers and academic researchers also hold very different knowledge about the species that reflects their equally different interests and values. For local fishers, *Aratus* are central to livelihood activities as the crab is collected for both subsistence and economic practices. Fishers' knowledge about *Aratus* therefore reflects their value for local livelihoods. For example, we learned from fishers that *Aratus* are attracted by making noises and beating mangrove branches against each other—a practice that is central to local harvesting practices. This is explained through some odor from mangrove trees that lures them and the property of *Aratus* showing curiosity. From academic knowledge, we can learn, for instance, of chemical elements perceived through olfaction that mediate interactions between animals, or that crabs perceive vibrations by sensory bristles at the articulations

of their legs. If we sum these inferences, we will build a more comprehensive account of *Aratus* and their relations with other entities and processes. But we may also make novel inferences by considering what we learn from both knowledge systems, for instance, that it is worth investigating whether the fishers could be exploring *Aratus*' sensory perceptions to capture them, by mimicking the vibrations produced by the crabs for communication during courtship (Maciel and Alves 2009). If we then consider, say, past pollution events that caused death of many of these crabs, about which we can learn from detailed information the fishers can provide, we could, as academic scientists, choose hypotheses worth testing about possible pollutants responsible for those events. The outcomes of this inquiry could help the communities not only understand but act in relation to environmental threats, e.g., demanding that local and state authorities make decisions regarding this problem that affects their livelihoods.

Overlapping taxonomies illustrate how diverse values can generate distinct bodies of knowledge and stimulate intercultural learning. While fishers and academics recognize the same crab species, their diverging interests and values guide them toward different knowledge about the crab that can be put into fruitful intercultural dialogue. At the same time, diverging interests and values can also lead to cases of diverging taxonomies in which the categories of local communities do not translate into the categories of academic researchers. Let us consider one such case, also investigated in Brazil (Ludwig et al. 2024; Robles-Piñeros 2021; Robles-Piñeros et al. 2020), in the farming communities of Coração de Maria and Retiro, places of cultural and economic struggle between the increasing dominance of industrial agricultural production and the livelihoods of peasant communities that hold rich agricultural traditions.

Jairo Robles-Piñeros found several overlaps and partialities between academic and local farmers' taxonomies, but we will concentrate here on an intriguing difference between their classificatory practices. This is the local classification of organisms in the *Pseudococcidae* (*Coccoidea*) family as fungi rather than *Hemiptera* insects, as in academic taxonomy. Known as scale insects, *Coccoidea* do not conform to the morphological pattern or behavior expected of insects. Adult females are almost always immobile, have soft bodies and no limbs, and permanently attach to the plant on which they are feeding, concealed underneath domed scales, extruding wax for protection. Males have legs and sometimes wings and resemble small flies, but they do not even exist in part of the *Coccoidea* species. Adult females call much more of the farmers' attention than males, because they feed on plants and thus show importance for farming practices. Going beyond behavioral and morphological features, also considering ecological interactions and agricultural significance, we can see why the farmers treat scale insects as fungi, based on their practices and interests. Scale insects attack crops like manioc in ways more similar to fungi rather than insects. The case of *Coccoidea* provides, thus, a window into the entanglement of ontologies, practices, and values. The farmers' expertise, goals, interests, and practices guide their classificatory attention to properties that scale insects share with fungi and justify why they do not treat *Coccoidea* as insects: behavioral (the females are sessile), morphological (they do not have an insect-like body), and ecological/agricultural (they interact with crops in ways similar to fungi). Meanwhile, for academic researchers *Coccoidea* are insects because they are phylogenetically related to other members of this class rather than fungi. Scientists and farmers use different practices, intertwined with different goals and interests, which generate different representational needs, leading to different ontologies.

But can farmers and academic taxonomists learn from their different categorization practices? Here it might seem more obvious how academic sciences could contribute to farming practices, as farmers can learn from scientists that scale insects are affected by insecticides rather than fungicides. It is important to notice, however, that this learning path is opened by agricultural practices dependent on pesticides that were introduced in the wake of the Green Revolution

and have been responsible for the decline of pineapple farming in Coração de Maria and Retiro (Robles-Piñeros 2021). Before the introduction of pesticides, local farmers used consortia of plants for controlling insects and decreasing harm to their harvests. When they were encouraged by agricultural technicians to “modernize” their practices and plant pineapples in monocultures, there was an increasing displacement of local agricultural practices, ultimately leading to increasing dependence on pesticides, an explosion of insect populations feeding on pineapples, and an evolution of pesticide resistance. This transition to “modernized” agriculture was one of the reasons for the decline of pineapple farming in Coração de Maria and Retiro. This is also related to how diverging values favor diverging taxonomies. In a community that prioritizes traditional farming methods, it is more useful to classify *Coccoidea* with fungi as they resemble fungi not only in their morphology and behavior but also in their ecological role in traditional farming systems. If industrialized monoculture farming is prioritized, however, it would be misleading to categorize *Coccoidea* with fungi rather than insects as they will be affected by insecticides rather than fungicides. Which taxonomy is more appropriate in a community depends on which practices it values.

Thus, seriously engaging with differences between knowledge systems uncovers cases in which different—and sometimes conflicting—value systems are generative of different taxonomic systems. Such cases stress the role of intercultural translation in dialogue and mutual learning between heterogeneous social actors, which gives salience, in turn, to the challenges involved in transdisciplinary and intercultural collaborations. If these challenges are taken to entail full incommensurability between knowledge systems, this will mean the impossibility of a translation that might allow for mutual intelligibility of different languages, cognitive experiences, knowledges, and symbolic universes, undermining any possibly fruitful interaction between heterogeneous social actors. Intercultural translation departs, however, from the idea that mutual understanding can take place, often assuming a deeply political nature as a way of struggling against the silencing of diverse cultures and knowledge systems through colonialism and globalization.

36.5 Concluding remarks

This chapter began by explaining the philosophical approach to the politics of knowledge on which the partial overlaps framework is grounded, showing how it emerges from a productive tension between epistemic diversity, as the basis for a criticism of epistemic paternalism that emphasizes collaboration between different knowledge systems, and a decolonial critique of both paternalism and transdisciplinarity that stresses tensions and differences that should not be neglected while searching for similarities and synergies. Partial overlaps in value commitments were then discussed by critically reflecting on claims that Indigenous and academic practices overlap in valuing sustainability. Finally, it was discussed how mutual learning between heterogeneous social actors with distinct value systems can take place based on both overlaps and their partialities.

The partial overlap framework provides a fruitful entry point for situating debates about science and values in complex empirical and intercultural landscapes. Contemporary philosophy of science emphasizes both the value-ladenness of science and the importance of diversity in scientific practices, raising the question of how different and sometimes conflicting values can be related to each other. This challenge is particularly pressing in transdisciplinary and intercultural research that involves actors with strikingly different values as well as different positions of power that commonly privilege the values of dominant actors. As shown in this chapter, the partial overlaps framework allows for nuanced interpretations of such complex landscapes by allowing for equal attention to similarities and differences between value systems. As such, the partial overlap approach challenges the idea that global science can be built on universal values that are shared by all affected actors without endorsing

a value relativism in which different forms of knowledge production are grounded in entirely incommensurable value systems.

Although this chapter follows the philosophical literature in arguing that values are not a problem to be eliminated from knowledge production, it does not mean to suggest that the roles of values are always innocent. On the contrary, discussions of both epistemic paternalism and extractivism show that dominant values can entail epistemic and social injustice, while also undermining transdisciplinary and intercultural collaboration, leading back to normative philosophical debates about both the epistemic and political legitimacy of different roles of values in research (de Melo-Martín 2024; Holman and Wilholt 2022).

References

- Alcoff, Linda M. 2022. "Extractivist Epistemologies." *Tapuya: Latin American Science, Technology and Society* 5 (1): 2127231. <https://doi.org/10.1080/25729861.2022.2127231>.
- Berlin, Brent. 1992. *Ethnobiological Classification*. Princeton University Press.
- Binns, Tony. 2014. "Dualistic and Unilinear Concepts of Development." In *The Companion to Development Studies*, edited by Vandana Desai and Rob Potter. Routledge.
- Boogaard, Birgit. 2021. "Epistemic Injustice in Agricultural Development: Critical Reflections on a Livestock Development Project in Rural Mozambique." *Knowledge Management for Development Journal* 16 (1): 28–54.
- Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Büscher, Bram, and Robert Fletcher. 2020. *The Conservation Revolution: Radical Ideas for Saving Nature beyond the Anthropocene*. Verso Books.
- Caniglia, Guido, and Federica Russo. 2024. "How Is Who: Evidence as Clues for Action in Participatory Sustainability Science and Public Health Research." *History and Philosophy of the Life Sciences* 46 (1): 4. <https://doi.org/10.1007/s40656-023-00603-5>.
- Caradonna, Jeremy L. 2022. *Sustainability: A History*. Oxford University Press.
- Chilisa, Bagele. 2019. *Indigenous Research Methodologies*. Sage Publications.
- Cornell, Sarah, Frans Berkhout, Willemijn Tuinstra, et al. 2013. "Opening Up Knowledge Systems for Better Responses to Global Environmental Change." *Environmental Science & Policy* 28: 60–70. <https://doi.org/10.1016/j.envsci.2012.11.008>.
- Cusicanqui, Silvia R. 2012. "Ch'ixinakax utxiwa: A Reflection on the Practices and Discourses of Decolonization." *South Atlantic Quarterly* 111 (1): 95–109. <https://doi.org/10.1215/00382876-1472612>.
- De La Rosa, Gabriela, Esther Milberg Muñiz, Charbel N. El-Hani, and David Ludwig. 2024. "Navigating between Promises and Realities of Transdisciplinary Research for Environmental Conservation." *Human Ecology* 52 (3): 697–708. <https://doi.org/10.1007/s10745-024-00520-6>.
- de Melo-Martín, Inmaculada. 2024. "Concerns about Contextual Values in Science and the Legitimate/Illegitimate Distinction." *Philosophy of Science* 91 (4): 851–68. <https://doi.org/10.1017/psa.2024.20>.
- Descola, Philippe. 2005. *Par-delà Nature et Culture*. Gallimard.
- Dewey, John. 1938. *Logic: The Theory of Inquiry*. Henry Holt and Company.
- Douglas, Heather E. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- El-Hani, Charbel N., and David Ludwig. 2025. "Intercultural Education as Dialogue between Knowledge Systems: Elements of a Theoretical Framework." *Science & Education* 34: 2797–844. <https://doi.org/10.1007/s11191-024-00525-z>.
- El-Hani, Charbel N., Luana Polisel, and David Ludwig. 2022. "Beyond the Divide between Indigenous and Academic Knowledge: Causal and Mechanistic Explanations in a Brazilian Fishing Community." *Studies in History and Philosophy of Science* 91: 296–306. <https://doi.org/10.1016/j.shpsa.2021.11.001>.
- Elkington, John. 2013. "Enter the Triple Bottom Line." In *The Triple Bottom Line*, edited by Adrian Henriques and Julie Richardson. Routledge.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Escobar, Arturo. 2020. *Pluriversal Politics: The Real and the Possible*. Duke University Press.
- Gomis, Alexis J. B., Manuel Guillén Parra, W. Michael Hoffman, and Robert E. McNulty. 2011. "Rethinking the Concept of Sustainability." *Business and Society Review* 116 (2): 171–91. <https://doi.org/10.1111/j.1467-8594.2011.00381.x>.

- Harding, Sandra. 2019. *Objectivity and Diversity: Another Logic of Scientific Research*. University of Chicago Press.
- Holden, Erling, Kristin Linnerud, and David Banister. 2014. "Sustainable Development: Our Common Future Revisited." *Global Environmental Change* 26: 130–9. <https://doi.org/10.1016/j.gloenvcha.2014.04.006>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Kimmerer, Robin W. 2012. "Searching for Synergy: Integrating Traditional and Scientific Ecological Knowledge in Environmental Science Education." *Journal of Environmental Studies and Sciences* 2 (4): 317–23. <https://doi.org/10.1007/s13412-012-0091-y>.
- Leonelli, Sabina. 2023. *Philosophy of Open Science*. Cambridge University Press.
- Lewis-Jones, Kay E. 2016. "'Useful to Us in Unknown Ways': Seed Conservation and the Quest for Novel Human-plant Relationships for the 21st Century." *Journal of Ethnobiology* 36 (1): 66–85. <https://doi.org/10.2993/0278-0771-36.1.66>.
- Ludwig, David. 2016a. "Overlapping Ontologies and Indigenous Knowledge. From Integration to Ontological Self-determination." *Studies in History and Philosophy of Science Part A* 59: 36–45. <https://doi.org/10.1016/j.shpsa.2016.06.002>.
- Ludwig, David. 2016b. "Ontological Choices and the Value-Free Ideal." *Erkenntnis* 81 (6): 1253–72. <https://doi.org/10.1007/s10670-015-9793-3>.
- Ludwig, David, and Birgit K. Boogaard. 2021. "Making Transdisciplinarity Work: An Epistemology of Inclusive Development and Innovation." In *The Politics of Knowledge in Inclusive Development and Innovation*, edited by David Ludwig, Birgit Boogaard, Phil Macnaghten, and Cees Leuwis. Routledge.
- Ludwig, David, and Charbel N. El-Hani. 2020. "Philosophy of Ethnobiology: Understanding Knowledge Integration and its Limitations." *Journal of Ethnobiology* 40 (1): 3–20. <https://doi.org/10.2993/0278-0771-40.1.3>.
- Ludwig, David, and Charbel N. El-Hani. 2025. *Transformative Transdisciplinarity: An Introduction to Community-Based Philosophy*. Oxford University Press.
- Ludwig, David, Daniel F. Banuoku, Birgit Boogaard, et al. 2024. "Southern Ontologies: Reorienting Agendas in Social Ontology." *Journal of Social Ontology* 10 (2). <https://doi.org/10.25365/jso-2024-7691>.
- Ludwig, David, and Stéphanie Ruphy. 2024. "Scientific Pluralism." *Stanford Encyclopedia of Philosophy*. Last Modified Fall 2024. <https://plato.stanford.edu/entries/scientific-pluralism/>.
- Maciel, Daniele C., and Ângelo G. C. Alves. 2009. "Conhecimentos e Práticas Locais Relacionados ao Aratu *Goniopsis cruentata* (Latreille, 1803) em Barra de Sirinhaém, Litoral Sul de Pernambuco, Brasil." *Biota Neotropica* 9 (4): 29–36. <https://doi.org/10.1590/S1676-06032009000400002>.
- Merton, Robert. 1979/1942. "The Normative Structure of Science." In *The Sociology of Science*. University of Chicago Press.
- Milberg Muñoz, Esther, David Ludwig, and Charbel N. El-Hani. 2025. "Research as a Mangrove: Emancipatory Science and the Messy Reality of Transdisciplinarity." *International Review of Qualitative Research* 18 (2): 237–59. <https://doi.org/10.1177/19408447241260446>.
- Moore, Jason W. 2015. *Capitalism in the Web of Life: Ecology and the Accumulation of Capital*. Verso Books.
- Reydon, Thomas A., and Marc Ereshefsky. 2022. "How to Incorporate Non-Epistemic Values into a Theory of Classification." *European Journal for Philosophy of Science* 12 (1): 4. <https://doi.org/10.1007/s13194-021-00438-6>.
- Robles-Piñeros, Jairo. 2021. "Etnoecologia, Formação de Professores de Ciências e Letramento Ecológico: Desenvolvendo um Perfil Culturalmente Sensível." PhD diss., Federal University of Bahia. <https://repositorio.ufba.br/bitstream/ri/34768/1/Tese%20Doutoral%20Jairo%20Robles%20Pi%C3%B1eros%20PPGEFHC.pdf>.
- Robles-Piñeros, Jairo, David Ludwig, Geilsa Costa Santos Baptista, and Adela Molina-Andrade. 2020. "Intercultural Science Education as a Trading Zone between Traditional and Academic Knowledge." *Studies in History and Philosophy of Biological and Biomedical Sciences* 84: 101337. <https://doi.org/10.1016/j.shpsc.2020.101337>.
- Rolin, Kristina. 2019. "The Epistemic Significance of Diversity." In *The Routledge Handbook of Social Epistemology*, edited by Miranda Fricker, Peter J. Graham, David Henderson, and Nikolaj J. L. L. Pedersen. Routledge.
- Rose, Deborah B. 2002. "Indigenous Ecologies and an Ethic of Connection." In *Global Ethics and Environment*, edited by Nicholas Low. Routledge.
- Smith, Linda T. 1999. *Decolonizing Methodologies: Research and Indigenous Peoples*. Zed Books and University of Otago Press.

- van de Poel, Ibo. 2015. "Conflicting Values in Design for Values." In *Handbook of Ethics, Values, and Technological Design*, edited by Jeroen van den Hoven, Pieter E. Vermaas, and Ibo van de Poel. Springer.
- van Kerkhoff, Lorrae, and Nicole A. Szlezák. 2016. "The Role of Innovative Global Institutions in Linking Knowledge and Action." *Proceedings of the National Academy of Sciences of the United States of America* 113 (17): 4603–8. <https://doi.org/10.1073/pnas.0900541107>.
- van Maurik Matuk, Fernanda. A., Bas Verschuuren, Piero Morseletto, et al. 2023. "Advancing Co-Production for Transformative Change by Synthesizing Guidance from Case Studies on the Sustainable Management and Governance of Natural Resources." *Environmental Science & Policy* 149: 103574. <https://doi.org/10.1016/j.envsci.2023.103574>.
- WCED. 1987. *Our Common Future: Report of the World Commission on Environment and Development*. UN Documents: Gathering a Body of Global Agreements.

TRAGEDY OR TRANSITION?

How Science and Values Matter for Climate Change Frames

Daniel Steel

37.1 Introduction

More than a half-century ago, Garrett [Hardin \(1968\)](#) argued that many environmental problems are a *tragedy of the commons* (ToC), in which collective catastrophe results from individuals separately acting to promote their self-interests. In Hardin's famous example, each shepherd earns more by adding sheep to the common pasture, but if all shepherds do this then the commons will be ruined, and all the shepherds will be left destitute. Climate change mitigation is often seen as having a similar economic structure ([Brennan 2009](#); [Li et al. 2021](#)). Climate change mitigation refers to actions taken to reduce atmospheric greenhouse gas concentrations by reducing emissions or enhancing carbon sinks. Replacing coal-burning power plants with solar farms and promoting the regrowth of forests are examples of climate change mitigation.

Characterizing climate change mitigation as a ToC is an example of a *frame* ([Patt 2017](#)). Frames set out basic assumptions that point toward key problems and solutions. Thus, framing climate change as a ToC assumes that, while climate change mitigation is collectively beneficial, cutting greenhouse gas emissions is costly for each country. Given these assumptions, the problem is that all countries have a strong incentive to free-ride on the mitigation of others. And if this is the problem, then solutions must discourage free-riding, for example, through an international agreement to penalize countries that shirk their fair share of mitigation ([Brennan 2009](#)).

While the ToC framing is common among economists ([Alestra et al. 2022](#); [Brennan 2009](#); [Lopez, Pastén, and Gutiérrez Cubillos 2022](#)), scientists who study technological systems often argue that climate change should be seen instead as an *energy transition* ([Mercure et al. 2021](#); [Newell 2021](#); [Patt 2017](#)). Unlike a ToC, energy transitions result mostly from decisions motivated by economic and strategic self-interest. For example, falling costs of electricity generated from wind and solar can make climate change mitigation in the power sector an economically superior option, especially for countries that would otherwise import fossil fuels ([Mercure et al. 2021](#)). The transition in Uruguay's power sector from fossil fuels to wind turbines is an oft-cited example of this phenomenon ([Bertoni and Messina 2023](#)). And China's ambition to become the world leader in renewable energy technologies and electric vehicles illustrates the role that geopolitical strategic interests can play in driving an energy transition ([Jaffe 2018](#)). According to the energy transition framing, transitioning from fossil fuels to renewable energy as quickly as possible is the key to solving climate change ([Sharpe 2023](#)). From this perspective, a binding international climate treaty is relatively

unimportant compared to falling prices and improved performance of renewable energy technologies (Patt 2015).

This chapter explains how value judgments as well as scientific findings are important when considering which framing of climate change should be adopted. Since frames inevitably simplify a complex situation and can shape research programs and government policies, judgments about the adequacy of a frame can depend in part on its moral effects (Ortmann and Veit 2023). Philosophers have pointed out ethical downsides of framing climate change as a ToC. For example, ToC frames entail that cutting greenhouse gas emissions is always a sacrifice for individual people and countries, and sacrifice frames, moreover, are often argued to be ineffective motivators (Prinzing 2023). Consequently, the ToC framing risks making failure to act on climate change a self-fulfilling prophecy (Kopeck 2017). Energy transition frames also confront ethical concerns. Some argue that priority for a rapid transition can be used as a cover for imposing harms on marginalized groups in the rush to mine materials, like lithium (Canelas and Carvalho 2023; Jerez, Garcés, and Torres 2021).

The purpose of this chapter is not to defend one climate change frame over another, but to make the point that both scientific evidence and ethics are relevant to which frame should be chosen. This discussion is organized into five sections as follows. Section 37.2 discusses the ToC and energy transition frames of climate change in greater depth, Section 37.3 discusses moral concerns relating to the two frames, Section 37.4 considers the relationship between climate frames and the role of values in science, and Section 37.5 concludes with reflections about the nature of frames.

37.2 Tragedy of the commons or energy transition?

A ToC, also known as an n -person prisoner's dilemma, is a game among multiple players who can choose to cooperate in or defect from efforts to secure a collective good. In Hardin's example, the collective good is a shared pasture where sheep can graze. Thus, shepherds cooperate when they refrain from grazing more sheep on the common pasture or defect when they add more sheep. For climate change mitigation, the collective good is an atmosphere with greenhouse gas concentrations suitable for human life, and a country cooperates when it reduces its greenhouse gas emissions or defects when it increases them. A ToC is defined by a combination of *collective benefits* and *individual costs*. Collective benefits mean that everyone is better off when all cooperate than when all defect, while individual costs mean each player is better off defecting no matter what the others do.

Arguments in favor of a ToC framing, therefore, need to make the case that climate change mitigation involves a combination of collective benefits and individual costs. In other words, they need to show: (1) Collective mitigation would generally make countries better off, but (2) each country is economically better off not mitigating, no matter the actions of other countries. The central argument for viewing climate change as a ToC arises from economic theory about public goods (Brennan 2009).

A public good is a benefit that all can enjoy, whether or not they have paid to maintain it. Environmental benefits, such as improved air quality or a stable climate, are public goods, in contrast with privately owned goods, like a house, where the owner has the right to exclude others. The argument, then, begins with the premise that climate change mitigation is a public good. Economic theory predicts that public goods will be underfunded in the absence of collective action requiring individuals to pay their share. That is because individuals who do not pay to support the public good can still enjoy its benefits, so free-riders are better off than those who contribute. Thus, Geoffrey Brennan writes, "decentralised action in relation to public goods involves incentives to free-ride that can only be satisfactorily overcome by explicitly collective action" (Brennan 2009, 310). Brennan conceives of collective action as a social contract, overseen by the government, that aims to protect public goods. In this contract, individuals give up their ability to free-ride on the condition that others are similarly restricted (Brennan 2009, 311). Given this reasoning, the fundamental problem of climate change is

the absence of a world government capable of enforcing an international treaty mandating emissions reductions. Brennan acknowledges that some may be moved by moral or other reasons to cut emissions at a cost to themselves, even without an international agreement. However, he suggests that unilateral mitigation will inevitably be modest and “of a largely token character” (Brennan 2009, 309).

However, scholars who approach climate change from the perspective of technological systems often claim that it should be framed as an *energy transition* rather than a ToC (Mercure et al. 2021; Newell 2021; Patt 2015). Energy transitions refer to shifts in primary fuel sources or technologies in an energy system (Sovacool 2016). Historical examples include the transition from wind to steam power in shipping (Geels 2002), and from horses to internal-combustion-engine-powered vehicles in road transport (Geels 2005). The energy transition framing, then, starts with the assumption that the persistence of a world economy powered by fossil fuels is incompatible with crucial climate targets, such as keeping mean global warming below 2°C (Patt 2015). From this perspective, the problem is to transition as rapidly as possible from fossil fuels to energy sources that do not emit greenhouse gases, and the solution, therefore, lies in the widespread uptake of renewable energy and associated technologies like lithium-ion batteries and electric vehicles.

Anthony Patt (2017) argues that energy transitions lack defining features of a ToC. Importantly, while energy transitions can generate collective benefits, they do not require the majority of decision-makers to act against their self-interests. To the contrary, the strategic and economic self-interests of governments and individuals usually play important roles in advancing energy transitions. For instance, starting in the eighteenth century, Great Britain subsidized the development of steamships to protect its position as the world’s dominant shipping and naval power (Geels 2002). Steamships were initially limited to niche uses, such as in-land waterways, but as the technology improved, the greater speed of steam-powered vessels, and their ability to stick to pre-determined schedules led to more widespread adoption (Geels 2002). Since energy transitions usually result from many actions motivated by strategic or economic interests, they are less prone to being stymied by free-riding. In fact, the steamship example illustrates that falling behind in an energy transition can carry risks of being at a competitive disadvantage.

Consider strategic interests in connection with renewable technologies. In 2023, about 80% of the manufacturing capacity for solar panels and lithium-ion batteries, along with about 60% of the world’s wind turbine manufacturing capacity, resided in China (IEA 2024). China is also the world’s largest producer of electric vehicles (IEA 2023). Government policies that subsidize Chinese companies to develop and manufacture renewable technologies are an important part of how all this happened (Qi et al. 2022). The motivations for China’s renewable subsidies appear to include a combination of difficulties in securing reliable, affordable oil imports to fuel its economic growth and a strategic vision of dominance in renewable technologies as a basis for geopolitical influence (Jaffe 2018). China’s role in the renewable energy transition, then, is analogous to Great Britain’s in the transition to steam power in shipping two centuries earlier. In both cases, national economic and strategic interests spurred a nation’s investments in new technologies that eventually upended incentives for others.

Moreover, unlike a ToC, decisions of actors in an energy transition are interdependent because actions of leaders can reduce costs and increase benefits to others of adopting new technologies. Since 2010, photovoltaic solar has gone from being 414% more expensive than fossil fuel-generated electricity to being 56% cheaper (IRENA 2024). Prices of onshore wind power have also flipped from being more to less expensive than fossil fuel alternatives, and these falling prices have coincided with a steadily increasing rate of installed renewable capacity (IRENA 2024). Prices of batteries have also consistently declined since the 1990s while their energy density has improved, thereby increasing the range of applications in which they can displace fossil fuels (Walter et al. 2023). All of these examples illustrate the experience curve, a well-known concept in technology diffusion according to which

costs fall and performance improves as a new technology's quantity of production increases (Nemet 2019). As a result of the experience curve, policies of one country, such as China's investments in renewable technologies, can bring down prices and improve performance, thereby making it more economically attractive for others to replace fossil fuels with renewables.

Energy transitions can drastically reduce pollution tied to a specific energy source. For instance, the transition from horses to gasoline and diesel-powered vehicles in road transport eliminated manure as an urban pollution problem (Patt 2015). While this transition did not occur because of public outcry about horse feces on city streets (Geels 2005), it got rid of the manure nonetheless. In the same way, a transition from fossil fuels could remove the source of the majority of the world's greenhouse gas emissions, even if most people are unwilling to undertake costly actions to mitigate climate change. However, as the transition from horse power to internal combustion engines illustrates, energy transitions often introduce new environmental problems. For a renewable energy transition to produce collective benefits, therefore, it is necessary that its environmental impacts be less harmful than those of fossil fuels. Given how environmentally damaging fossil fuels are, this is a fairly low bar to clear (Patt 2015). After all, fossil fuels have many other adverse environmental and public health impacts besides climate change, including particulate air pollution (Karlsson, Alfredsson, and Westling 2020). Nevertheless, environmental risks associated with renewable energy technologies, such as lithium mining, are a genuine concern. We consider this issue in the next section.

37.3 Values and climate change frames

The discussion so far has focused on disputes about how to frame climate change among social scientists in such fields as economics, political science, science and technology studies, and foreign affairs. I now turn to discussions of this issue from an explicitly ethical perspective. The main ethical concern about ToC framings is that they discourage action on climate change and thus risk a self-fulfilling prophecy of environmental catastrophe (Kopec 2017; Ortmann and Veit 2023; Prinzing 2023). In the case of energy transition frames, a number of scholars worry that installing renewable energy at maximum speed would run roughshod over the rights of poor and historically marginalized people (Canelas and Carvalho 2023). This section examines these two lines of concern.

Ethical concerns about representing climate change mitigation as a ToC hinge on a basic difference between physical and human sciences. The behavior of an atom is unaffected by whether or not humans choose to represent it by means of the Rutherford model, in which negatively charged electrons circulate around neutrons and protons that form the atom's nucleus. By contrast, humans can become aware of the concepts, hypotheses, and theories proposed by social scientists, and this awareness may affect their behavior (Hacking 1999). Thus, several authors argue that ToC framings can undermine effective action on climate change by promoting pessimism (Ortmann and Veit 2023), depicting pro-environmental behaviors as sacrifices (Prinzing 2023), and generating a self-fulfilling prophecy (Kopec 2017).

Framing climate change as a ToC can encourage pessimism because it entails that a world government is needed for effective mitigation. Without a world government with the power to prevent free-riding, a ToC framing predicts that mitigation will be limited to token actions (Brennan 2009). Of course, no world government with the power to compel countries to reduce emissions exists, and it seems extremely unlikely that one will be created any time soon. Indeed, efforts to broker an international agreement requiring countries to cut emissions have been largely unsuccessful. Since 1992, international climate negotiations have been carried out under the auspices of the United Nations Framework Convention on Climate Change (UNFCCC). The UNFCCC led to the Kyoto Protocol, which went into effect in 2005 and involved an agreement of 37 industrialized nations to reduce their emissions by 5% in comparison to 1990 levels. However, neither the United States nor China, the

world's two largest greenhouse gas emitters, was among those 37 countries. The Kyoto Protocol was replaced by the Paris Agreement in 2021. The Paris Agreement, itself adopted in 2015, affirms the aim of keeping mean global warming well below 2.0°C, but unlike the Kyoto Protocol, it is not an agreement to reduce emissions. Instead, participating states agree to propose nationally determined contributions toward addressing climate change and to provide regular updates on their progress. The amount of mitigation promised in nationally determined contributions, as well as their implementation, is entirely voluntary.

Thus, if climate change mitigation is a ToC, then there is little reason for optimism that it can be addressed effectively. Without a binding international climate agreement, ToC predicts that free-riding will undermine any efforts at collective action (Brennan 2009). Given the ToC framing, then, collective action is not possible under the current circumstances, while unilateral reductions of greenhouse gas emissions are futile and irrational (Ortmann and Veit 2023). It would be difficult to imagine a more pessimistic and discouraging view of climate change than this!

Some argue that characterizing mitigation as a sacrifice for the common good is not a good way to motivate action (Prinzing 2023). A “self-sacrifice framing” portrays pro-environmental behaviors as “personal sacrifices—bad for the person performing them (compared with less environmentally friendly alternatives) but good for the planet and/or other people, including future generations” (Prinzing 2023, 3). Characterizing climate change as a ToC is a self-sacrifice framing in this sense because it portrays cutting greenhouse gas emissions as acting against one's self-interests in order to secure a stable, habitable climate for all. According to Michael Prinzing, self-sacrifice framings are poor motivators because people tend to be highly averse to giving up advantages that they already possess (Prinzing 2023). A better approach, Prinzing suggests, is to emphasize the personal and local benefits of pro-environmental behaviors. For example, replacing a gas-fired boiler with a heat pump in your house can reduce your utility bill while improving indoor air quality. At a community level, transitioning to renewables and electric vehicles would reduce local air pollution. Prinzing also suggests that less materialistic and more environmentally conscious lifestyles often lead to greater personal well-being (Prinzing 2023). Rather than framing climate action as a sacrifice, therefore, Prinzing suggests that we “spread the word that going green is good for you!” (Prinzing 2023, 10).

Finally, Matthew Kopec (2017) argues that a ToC framing can lead to a self-fulfilling prophecy of failed attempts to address climate change. Kopec suggests that an economic model is prone to becoming a self-fulfilling prophecy when it appears scientifically compelling and belief in the model causes people to behave in ways that make the model more closely approximate reality. Framing climate change as a ToC arguably fits this description. The ToC frame of climate change seems to directly follow from well-established economic theorizing about public goods (Brennan 2009). And if political leaders and many members of the general public view climate change as a ToC, then they are likely to see mitigation as a futile sacrifice. And in that case, they are unlikely to support policies that would reduce greenhouse gas emissions. The failure of countries to seriously pursue climate change mitigation would then be taken as further evidence for a ToC framing, thus further discouraging mitigation, and so on. The potential for the ToC framing to generate a disastrous self-fulfilling prophecy, Kopec suggests, is a reason to consider alternative frames.

Framing climate change as an energy transition is a much better option when considered in connection with the above worries. Energy transitions do not need a global government to enforce them and do not demand mass sacrifice for a greater good. Furthermore, if belief in the inevitability of a transition from fossil fuels to renewables became a self-fulfilling prophecy, the results, such as more rapid climate change mitigation and reduced air pollution, would be broadly positive. Nevertheless, ethical concerns can arise for energy transition frames as well.

The chief moral concern with framing climate change as an energy transition is that it encourages a zeal to prioritize the rapid uptake of renewable energy above all else. While renewable energy

technologies are less harmful than fossil fuels, they can still have significant human health and environmental impacts. Consider these points in connection with lithium mining. As noted in the previous section, falling prices of batteries are a crucial component of a transition away from fossil fuels, and lithium-based batteries currently dominate the market (Walter et al. 2023). Utility-scale electricity storage is necessary for renewables to fully displace fossil fuels from the power sector, and batteries are currently the leading option for this purpose. Low-cost batteries with high energy density are also needed to drive fossil fuels out of transport.

Lithium batteries include other materials besides lithium itself, prominently including cobalt, nickel, and magnesium. Pressure to produce lithium batteries as cheaply as possible can encourage cutting costs at the expense of worker safety and environmental protection when mining these materials. A recent review examined research on the human health impacts of mining associated with lithium batteries (Brown et al. 2024). This review found studies providing evidence of toxicity related to cobalt, nickel, and magnesium mining. Although lithium is known to be potentially toxic to humans, and there have been cases of toxicity associated with lithium mining reported in the media, no academic studies on the health effects of lithium mining were found in the review (Brown et al. 2024). In sum, prioritizing cheap lithium batteries can exacerbate human health risks, although these risks are impossible to quantify given current research gaps.

Consequently, several authors argue that by excessively prioritizing the rapid uptake of renewable energy technologies, an energy transition framing can be harmful (Canelas and Carvalho 2023; Jerez, Garcés, and Torres 2021). The energy transition framing, it is argued, can serve as a cover for mining corporations and governments to implement projects with scant regard for their adverse impacts on local populations and ecosystems. This type of scenario is especially worrisome when the populations in question are politically marginalized. For example, in the “Lithium Triangle,” an Andean region in South America that overlaps Chile, Argentina, and Bolivia, containing the world’s largest lithium deposits, impacted populations prominently include Indigenous Peoples (Liu and Agusdinata 2020). The rights of Indigenous Peoples were also a central issue in the case of the Thacker Pass lithium mine in the US state of Nevada (Rodeiro 2023). In Europe, proposed lithium mining projects have sometimes generated backlash along rural versus urban lines (Canelas and Carvalho 2023).

From the perspective of critics, then, an energy transition framing risks perpetuating economic and political forces that generate harmful environmental and human health effects. While specific technologies may be linked to particular types of pollution, environmental injustices more generally result from economic and political systems that treat certain groups of people as expendable in the name of greater profits and power for others (Givens, Huang, and Jorgenson 2019). According to this view, the fact that fossil fuels are worse than renewables is not sufficient justification for an energy transition framing. A better alternative, from this perspective, would be a *just transition* framing that emphasizes the need to both transition from fossil fuels and promote greater economic and political fairness (Wang and Lo 2021).

37.4 Values in science and moral imagination

The arguments in the previous section all presumed that ethical value judgments are relevant to which framing of climate change should be adopted. But aren’t climate change frames scientific hypotheses to be accepted or rejected solely on the basis of evidence? This question leads us to debates about values and science.

Everyone agrees that science is linked to important social and moral values in some respects. For example, there should be ethical constraints on scientific research involving human subjects, and values matter when deciding what actions or policies to implement in light of scientific findings. As a result, philosophical debates about values and science usually focus on what one might

think of as “internal” aspects of science, in which scientific claims are assessed in light of data. Estimating the effect of increased greenhouse gas concentrations on global mean temperature would be an example of a topic that is internal to science. The *value-free ideal* claims that, to prevent bias, values should have no role in this internal domain (Betz 2013). Given this outlook, scientists should do their best to minimize the impact of their own values on their assessments of what claims the data do and do not support. In contrast, the *argument from inductive risk* aims to show that values are relevant even for activities internal to science (Douglas 2009). According to the argument from inductive risk, scientists must make decisions about what claims to accept, and this entails a judgment about how much evidence is sufficient. If significant harm would result from asserting something that turns out to be false, the argument goes, ethical values can be relevant to deciding how much evidence is enough. The greater the harm that would result from mistakenly accepting the claim, the stronger the evidence needs to be. Heather Douglas (2009) refers to this as an indirect role of values and distinguishes it from a direct role in which one simply accepts claims that agree with one’s values. According to Douglas, direct roles of values should be prohibited in the internal areas of science, but values in an indirect role can permeate science without compromising its objectivity.

While debates about the value-free ideal raise important questions concerning the nature of science, bias, and objectivity, their relevance to climate change frames is questionable. For starters, frames exist in the borderlands of science and society, rather than being internal to science itself. Frames are general interpretations of an issue that emerge at the intersection of science, government, stakeholders, and the general public (Koon, Hawkins, and Mayhew 2016). Frames play an important role in communicating the meaning and practical relevance of scientific findings, so they often involve a combination of value judgments and empirical claims that, taken together, suggest policy recommendations (Elliott 2017). Frames, then, are a type of “mixed claim,” in which value judgments are unavoidably embedded in empirical generalizations (Alexandrova 2018). The above points are illustrated by the ToC framing. It implies that public goods should be maintained (value judgment) but that doing so requires a collective action to discourage free-riding (empirical claim), and thus that an international climate treaty should be sought (policy recommendation). Since frames are not internal to science, even a person who accepts the value-free ideal could agree that they involve value judgments.

In addition, the argument from inductive risk is not well-suited for understanding the primary ethical concern raised about climate change frames, namely, self-fulfilling prophecies. The argument from inductive risk emphasizes harms that could result from asserting something that turns out to be false. For example, if scientists mistakenly declare a food additive to be safe, then people may be sickened as a result of eating it. In this example, scientists’ claims about the safety of the additive can change what people eat, but not whether the additive is in fact safe or toxic. That is very different from a self-fulfilling prophecy, in which asserting a claim ultimately makes that claim true when it could have otherwise been false. For instance, if you tell yourself that you are no good at math, then your math scores will likely be worse than if you had a more positive attitude. In the same way, critics worry that, by portraying free-riding as rational in the absence of collective action, ToC frames discourage countries from reducing their emissions, which is then interpreted as evidence for the accuracy of the ToC framing. Likewise, critics worry that energy transition frames, by placing such a high priority on the rapid expansion of renewable technologies, unnecessarily encourage an unjust transition from fossil fuels. The concerns here aren’t about asserting something that turns out to be false, but instead that asserting something bad will make the bad thing come true. So, it seems that another approach to values and science may be called for here.

One option can be found in Matthew Brown’s *Science and Moral Imagination* (2020). Brown argues that debates about values in science should not be limited to minimum boundaries intended

to prevent the harmful effects of values. In addition, Brown argues that it is necessary to specify a positive ideal that science should strive for. For this, [Brown \(2020\)](#) proposes an “ideal of moral imagination,” according to which:

Scientists should recognize the contingencies in their work as unforced choices, discover morally salient aspects of the situation they are deciding, empathetically recognize and understand the legitimate stakeholders, imaginatively construct and explore possible options, and exercise fair and warranted value judgment in order to guide those decisions. (21)

Consider how this ideal applies to the case we have been discussing. Framing climate change is an example of a contingency, because there are multiple options with differing empirical and ethical pros and cons. So, rather than taking one frame (say, ToC) as a given, the first step is to recognize that a choice among frames exists. The next step is to consider morally salient aspects of that choice. The previous section identified the potential for harmful self-fulfilling prophecies as a key issue here. The ideal of moral imagination then recommends working with legitimate stakeholders to explore possible options and choose among them.

Brown’s ideal of moral imagination is intended as a broad guideline, not an algorithm. Judgment is involved, for example, in deciding who is a “legitimate stakeholder” and what counts as “warranted value judgment.” The main point here, however, is to encourage creative exploration of possible framings of climate change, and to emphasize that this is a choice with moral consequences. Contingency is especially important to remember in connection with framing. Widely accepted framings can be taken for granted, which can in turn guide science and policy in ways that may have negative consequences. So, an imaginative reevaluation of climate change frames is an important exercise.

37.5 Conclusions

By providing starting assumptions that suggest key problems and solutions, frames guide how people understand important topics like climate change. Climate change also illustrates that framing usually involves a choice. Frames are simplifications that make a complex reality easier to grasp, so multiple frames that highlight different aspects of a complex situation can almost always be proposed. The old adage of blind men and the elephant may be helpful to convey this idea. In the story, a group of blind men who have never encountered an elephant attempt to describe one by touch, producing very different results depending on which part of the elephant they come in contact with. For example, one touches a leg and says an elephant is a pillar, while another touches an ear and says that an elephant is flat and supple like a cloth. In a similar manner, different frames tend to focus on distinct aspects of a broad topic and often each has some grain of truth.

This chapter has focused on two frames, one in which climate change mitigation is seen as a ToC and the other that emphasizes an energy transition from fossil fuels to renewables. These are not the only possibilities. A just transition frame was also mentioned, and other frames have also been discussed ([Steel, Mintz-Woo, and DesRoches 2023](#)). When assessing a climate change frame, it is important to consider how well it approximates central factual details. For example, the emergence of solar and wind power that can outcompete fossil fuels on price challenges the ToC’s core assumption that mitigation is always individually costly. However, value judgments also matter when assessing frames. Since frames can fundamentally shape the processes they describe, self-fulfilling prophecies are a major ethical concern. The future we would like to create and the facts as they are now, therefore, both matter for decisions about which climate change frame to adopt.

References

- Alestra, Claire, Gilbert Cette, Valérie Chouard, and Rémy Lecat. 2022. "Growth Impact of Climate Change and Response Policies: The Advanced Climate Change Long-Term (ACCL) Model." *Journal of Policy Modeling* 44 (1): 96–112. <https://doi.org/10.1016/j.jpolmod.2021.10.001>.
- Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *The British Journal for the Philosophy of Science* 69 (2): 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Bertoni, Reto, and Pablo Messina. 2023. "The 'Wind Revolution' in Uruguay and the Role of the Public Sector in Guiding Energy Transitions." In *Energy Transitions in Latin America*. Sustainable Development Goals Series, edited by Lira Luz Benites Lazaro and Esteban Serrani. Springer.
- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Brennan, Geoffrey. 2009. "Climate Change: A Rational Choice Politics View." *Australian Journal of Agricultural and Resource Economics* 53: 309–26. <https://doi.org/10.1111/j.1467-8489.2009.00457.x>.
- Brown, Connor W., Charlotte E. Goldfine, Lao-Tzu Allan-Blitz, and Timothy B. Erickson. 2024. "Occupational, Environmental, and Toxicological Health Risks of Mining Metals for Lithium-Ion Batteries: A Narrative Review of the PubMed Database." *Journal of Occupational Medicine and Toxicology* 19: 35. <https://doi.org/10.1186/s12995-024-00433-6>.
- Brown, Matthew. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. University of Pittsburgh Press.
- Canelas, Joana, and António Carvalho. 2023. "The Dark Side of the Energy Transition: Extractivist Violence, Energy (In)justice and Lithium Mining in Portugal." *Energy Research & Social Science* 100: 103096. <https://doi.org/10.1016/j.erss.2023.103096>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Geels, Frank. 2002. "Technological Transitions as Evolutionary Reconfiguration Processes: A Multi-Level Perspective and a Case-Study." *Research Policy* 31 (8–9): 1257–74. [https://doi.org/10.1016/S0048-7333\(02\)00062-8](https://doi.org/10.1016/S0048-7333(02)00062-8).
- Geels, Frank. 2005. "The Dynamics of Transitions in Socio-Technical Systems: A Multi-Level Analysis of the Transition Pathway from Horse-Drawn Carriages to Automobiles (1860–1930)." *Technology Analysis & Strategic Management* 17 (4): 445–76. <https://doi.org/10.1080/09537320500357319>.
- Givens, Jennifer E., Xiaorui Huang, and Andrew K. Jorgenson. 2019. "Ecologically Unequal Exchange: A Theory of Global Environmental Injustice." *Sociology Compass* 13: e12693. <https://doi.org/10.1111/soc4.12693>.
- Hacking, Ian. 1999. *The Social Construction of What?* Harvard University Press.
- Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162: 1243–8. <https://doi.org/10.1126/science.162.3859.1243>.
- IEA. 2023. "Global EV Outlook 2023." <https://www.iea.org/reports/global-ev-outlook-2023>.
- IEA. 2024. "Advancing Clean Technology Manufacturing." <https://www.iea.org/reports/advancing-clean-technology-manufacturing>.
- IRENA. 2024. "Renewable Power Generation Costs in 2023." International Renewable Energy Agency. <https://www.irena.org/Publications/2024/Sep/Renewable-Power-Generation-Costs-in-2023>.
- Jerez, Bárbara, Ingrid Garcés, and Robinson Torres. 2021. "Lithium Extractivism and Water Injustices in the Salar de Atacama, Chile: The Colonial Shadow of Green Electromobility." *Political Geography* 87: 102382. <https://doi.org/10.1016/j.polgeo.2021.102382>.
- Jaffe, Amy Myers. 2018. "Green Giant: Renewable Energy and Chinese Power." *Foreign Affairs* 97 (2): 83–93. <https://www.jstor.org/stable/44822083>.
- Karlsson, Mikael, Eva Alfredsson, and Nils Westling. 2020. "Climate Policy Co-Benefits: A Review." *Climate Policy* 20 (3): 292–316. <https://doi.org/10.1080/14693062.2020.1724070>.
- Koon, Adam, Benjamin Hawkins, and Susannah Mayhew. 2016. "Framing and the Health Policy Process: A Scoping Review." *Health Policy and Planning* 31 (6): 801–16. <https://doi.org/10.1093/heapol/czv128>.
- Kopec, Matthew. 2017. "Game Theory and the Self-Fulfilling Climate Tragedy." *Environmental Values* 26 (2): 203–21. <http://www.jstor.org/stable/44202482>.
- Li, Yang, David K. Sewell, Saam Saber, Daniel B. Shank, and Yoshihisa Kashima. 2021. "The Climate Commons Dilemma: How Can Humanity Solve the Commons Dilemma for the Global Climate Commons?" *Climatic Change* 164: 4. <https://doi.org/10.1007/s10584-021-02989-2>.

- Liu, Wenjuan, and Datu Agusdinata. 2020. "Interdependencies of Lithium Mining and Communities Sustainability in Salar de Atacama, Chile." *Journal of Cleaner Production* 260: 120838. <https://doi.org/10.1016/j.jclepro.2020.120838>.
- Lopez, Ramon, Roberto Pastén, and Pablo Gutiérrez Cubillos. 2022. "Climate Change in Times of Economic Uncertainty: A Perverse Tragedy of the Commons?" *Economic Analysis and Policy* 75: 209–25. <https://doi.org/10.1016/j.eap.2022.05.005>.
- Mercure, Jean-François, Pablo Salas, Pim Vercoulen, et al. 2021. "Reframing Incentives for Climate Policy Action." *Nature Energy* 6: 1133–43. <https://doi.org/10.1038/s41560-021-00934-2>.
- Nemet, Gregory. 2019. *How Solar Energy Became Cheap: A Model for Low-Carbon Innovation*. Routledge.
- Newell, Peter. 2021. *Power Shift: The Global Political Economy of Energy Transitions*. Cambridge University Press.
- Ortmann, Jakob, and Walter Veit. 2023. "Theory Roulette: Choosing that Climate Change Is Not a Tragedy of the Commons." *Environmental Values* 32 (1): 65–89. <https://doi.org/10.3197/096327122X16452897197784>.
- Patt, Anthony. 2015. *Transforming Energy: Solving Climate Change with Technology Policy*. Cambridge University Press.
- Patt, Anthony. 2017. "Beyond the Tragedy of the Commons: Reframing Effective Climate Change Governance." *Energy Research & Social Science* 34: 1–3. <https://doi.org/10.1016/j.erss.2017.05.023>.
- Prinzing, Michael. 2023. "Going Green Is Good for You: Why We Need to Change the Way We Think about Pro-Environmental Behavior." *Ethics, Policy & Environment* 26 (1): 1–18. <https://doi.org/10.1080/21550085.2020.1848192>.
- Qi, Xiaoyan, Yanshan Guo, Pibin Guo, Xilong Yao, and Xiuli Liu. 2022. "Do Subsidies and R&D Investment Boost Energy Transition Performance? Evidence from Chinese Renewable Energy Firms." *Energy Policy* 164: 112909. <https://doi.org/10.1016/j.enpol.2022.112909>.
- Rodeiro, Manuel. 2023. "Mining Thacker Pass: Environmental Justice and the Demands of Green Energy." *Environmental Justice* 16 (2): 91–95. <https://doi.org/10.1089/env.2021.0088>.
- Sharpe, Simon. 2023. *Five Times Faster: Rethinking the Science, Economics, and Diplomacy of Climate Change*. Cambridge University Press.
- Sovacool, Benjamin. 2016. "How Long Will It Take? Conceptualizing the Temporal Dynamics of Energy Transitions." *Energy Research & Social Science* 13: 202–15. <https://doi.org/10.1016/j.erss.2015.12.020>.
- Steel, Daniel, Kian Mintz-Woo, and C. Tyler DesRoches. 2023. "Collapse, Social Tipping Dynamics, and Framing Climate Change." *Politics, Philosophy & Economics* 23 (3): 230–51. <https://doi.org/10.1177/1470594X231196432>.
- Walter, Daan, Kingsmill Bond, Sam Butler-Sloss, Laurens Speelman, Yuki Numata, and Will Atkinson. 2023. "X-Change: Batteries." Rocky Mountain Institute. <https://rmi.org/insight/x-change-batteries/>.
- Wang, Xinxin, and Kevin Lo. 2021. "Just Transition: A Conceptual Review." *Energy Research & Social Science* 82: 102291. <https://doi.org/10.1016/j.erss.2021.102291>.

38

VALUES IN SCIENCE COMMUNICATION MODELS

T.Y. Branch

38.1 Introduction

Public science communication serves as a crucial bridge between scientific experts and non-experts, structuring interactions and defining expectations. This strategic role has prompted philosophers of science to explore its impact on folk epistemic notions of science and public trust in science. Investigating these themes through the lens of science communication requires careful consideration of the diverse strategies employed and their characteristics. This chapter traces the evolution and establishment of twentieth- and twenty-first-century forms of science communication. It differentiates features about them, such as how they attribute epistemic authority, direct communication, and—relevant for those interested in the discourse on values in science—the type of values they solicit and propagate.

Science communication is not a synonym for public awareness of science, public understanding of science, scientific culture, or science literacy (Burns, John O'Connor, and Stockmayer 2003). For the purposes of this chapter, I define science communication as the sharing of information aimed at fostering awareness and knowledge of science, evoking enjoyment or emotional responses, and stimulating interest. These strategies seek to nurture opinions and attitudes toward science to promote an understanding that integrates its culture and knowledge into broader society (Bryant 2003, 357). This definition is intentionally expansive, encompassing various forms of science communication—conveyed by scientific experts or non-scientist communicators, across different media and through different forms. The approach will enable tracking the evolution of models for science communication and show how norms for values in science are expressed through them.

The chapter examines Western science communication from the mid-twentieth to the early twenty-first century for two primary reasons. First, while the history of science communication can be traced back hundreds of years (Bucchi 2008), the field has undergone significant development over the past 50 years (Gascoigne et al. 2010), with most research focused on the West (Finlay et al. 2021). This body of work offers rich analyses of the evolution of science communication models (Hu 2024; Jamieson 2017), which will be central to my discussion. Second, following World War II, the United States and the United Kingdom exerted an outsized influence over global perspectives on the appropriate relationship between science and society,¹ particularly by using science communication to cultivate a knowledgeable public that supports science (Halpern and Elliott 2022). The design and implementation of science communication in this post-war context laid the groundwork for contemporary practices in the twenty-first century.

I will begin by situating the infamous deficit model of science communication within the mid-twentieth-century Social Contract for Science (SCS) described by [Branch and Douglas \(2024\)](#). The deficit model, which dominated from the 1950s to the 1980s ([Gross 1994](#); [Stocklmayer 2013](#)), continues to serve as the default framework in many sub-disciplines of science ([Trench and Junker 2001](#)). From there, I will briefly transition through the dialogue model of science communication to arrive at the public participation model, which introduces new avenues for public values to enter science. The chapter will conclude with an examination of participant-led research, illustrating how social values can be woven throughout science and communication in ways that diverge from earlier models. By illustrating how these various forms of science communication reflect norms for values in science, I aim to provide philosophers with new ways to explore the value management challenge through a three-model framework for understanding how publics encounter science.

38.2 The deficit model of science communication

The mid-twentieth-century SCS emerged after World War II. It scaffolded a relationship between science and society by synthesizing guidelines, norms, and ideals that had been proposed separately in earlier decades. Yet, these only coalesced into an influential blueprint for strengthening science in the West during the Cold War ([Douglas and Branch 2024](#)). The mid-twentieth-century SCS also introduced normative expectations for how research should be conducted through the Value-Free Ideal (VFI) for science, positing that personal, social, public, and political (i.e., non-epistemic values) should not influence the evaluation of evidence or the context of justification.² The SCS and the VFI isolated scientists doing basic research (as opposed to applied science) from societal values, which mid-twentieth-century science communication sought to capture and codify for the public, resulting in the deficit model.

The deficit model relies on two core assumptions: First, that scientists are the sole epistemic authority on science, and second, that the public—a homogenous entity—is ignorant of science. The public's insufficient knowledge of science is presumably remedied through increased exposure to the empirical findings discovered by science ([Ahteensuu 2012](#); [Aydeniz and Southerland 2012](#); [Bodmer 1987](#); [Bucchi 2008](#); [Suldovsky 2016](#)). In this model, scientists disseminate information unidirectionally (sometimes through intermediaries) to a passive audience, who are thus expected to improve their science literacy ([Durant, Evans, and Thomas 1989](#); [Gross 1994](#); [Miller 1983](#); [Moser and Dilling 2011](#)). In so doing, they are envisioned to develop a positive impression of science ([Sturgis and Allum 2004](#)). This approach reinforces the authority and isolation of basic scientists prescribed by the SCS, encourages ongoing public funding for science, and theoretically equips the next generation of scientists with foundational knowledge to fuel the linear model for science ([Branch and Douglas 2024](#)).

The deficit model also reinforces the VFI with respect to value norms. The deficit model stresses communicating the empirical findings of science or knowledge from the basic sciences. Although this knowledge is not explicitly specified as being value-free, the deficit model allows for it to be conveyed as such, by not expecting scientists to disclose values in their science or consider how the public might respond to the science, other than favorably. Furthermore, the model's unidirectional design prevents public values from reaching scientists. In essence, the deficit model's exclusive attribution of epistemic authority to scientists, portrayal of the public as an uninformed mass, unidirectional flow of information, and focus on basic science collectively support the (presumably possible) pursuit of science separate from non-epistemic values.

Support for the deficit model began to wane in the mid-1980s under a barrage of criticism. Detractors have pointed to its excessive optimism ([Cortassa 2016](#)) and its ineffectiveness at enhancing science literacy as reasons to abandon it ([Holland 2007](#); [Nisbet and Mooney 2007](#)). The model's overly simplistic portrayal of scientists, the public, and the complexities of science literacy has

further eroded its appeal (Hansen et al. 2003; Sturgis and Allum 2004). Moreover, the model's unfair characterization of those opposed to science as inherently deficient or uninformed (Priest 2001) has also fueled its decline.³ Despite these criticisms, the model is still adopted in the twenty-first century (Besley and Tanner 2011; Davies 2008), and its persistence should be attributed, at least in part, to the SCS which continues to undergird scientific communities and the institutions that support them.

38.3 The rise of the public participation model

The shift away from the deficit model included a pivot from promoting science literacy to prioritizing public attitudes toward science (Bauer 2008). In this new paradigm, generating positive impressions of science became the central aim as opposed to a byproduct of communicating the findings of science. The hope was that, by continuing to improve public enthusiasm toward science, greater interest and respect for science would follow.

At the onset of the 1990s, the dialogue model emerged to improve public attitudes toward science. The model encourages people to ask questions about science in order to improve their impressions of science (van der Sanden and Meijman 2008). By inviting inquiries, the model created avenues for bidirectional engagement and an opportunity for public values to enter scientific discourse. Yet, as scientists remained satisfied with responding to questions, showing little inclination to explore the public's expertise, the model continued to uphold the prevailing authority of science. This leads to the model being labeled "pragmatic" because it insists that only scientists are responsible for knowledge production (Bucchi 2008). The model is also described as "constructivist" (Trench 2008) because it appreciates how scientific information is absorbed in the context of social knowledge frameworks, including folk epistemologies of science. Most importantly, it reimagines the notion of a monolithic public by recognizing a tapestry of diverse *publics* instead.⁴ This perspective embraces a variety of groups, including non-experts, extra-academic specialists, and stakeholders, each shaped by their unique social milieu, epistemic perspective, attitudes, and values, but not as epistemic contributors to science.

The 1990s saw the rise of communication models that encouraged greater inclusion of publics throughout science. They afforded publics the opportunity to share expertise (Goldenberg 2016; Wynne 1992) and, unlike with other models, enabled the immediate integration of public values into the scientific process. These new endeavors blurred the lines between basic and applied science by situating scientific inquiry within socially relevant contexts. Research was being pursued to address specific questions and also in response to public interest, not merely for the sake of scientific curiosity. This shift facilitated an exchange of information and skills between scientists and publics, giving rise to what is now known as the public participation model of science communication. This model can be instantiated in different ways, but what unifies this account is its bidirectional communication and opportunities for collaboration. Under the deficit model, the public received the findings of science; the dialogue model permitted publics to ask questions about science, and public participation offered them the chance to contribute to science.

The manner in which publics are invited to participate varies significantly depending on the type of research project (Strasser et al. 2019). In some cases, individuals are asked to contribute material resources, such as personal computers, for making scientific calculations. This is exemplified by initiatives like SETI@home. In other instances, publics take on roles as data collectors, as seen with amateur naturalists participating in eBird or residents mapping air pollution through City Sense. Publics can also contribute a diverse array of data by sharing observations, symptoms, and experiences in medical studies such as PatientsLikeMe. In other scenarios, involvement extends to influencing the very design of experiments, such as advocating for new categories of galaxies in projects like Galaxy Zoo. Public participation not only yields pragmatic benefits but also enhances the social impact of academic research by increasing the representation

of underrepresented groups within scientific communities. It actively engages citizens and stakeholders in the governance of science and technology while fostering interdisciplinary collaboration (Koskinen 2022).

Some forms of participation offer more substantial engagement than others, such as donating computing power compared to developing new classification structures for galaxies. Independent of how publics participate in science, under the public participation model, they have a greater chance for their values to directly influence science than in previous models. This is in part because even when following a protocol laid out by scientists, publics will inevitably face uncertainty too—for example, on how to classify an ambiguous galaxy. Here, publics, like scientists, are compelled to weigh the consequences of making false-positive and false-negative errors. This is where non-epistemic values will inevitably play a part in evaluating evidence. Therefore, like scientists, public participants face inductive risks and will have to balance different values. Philosophers of science have come to a consensus that these decisions ought to include non-epistemic values (Brown 2024; Holman and Wilholt 2022), so how publics adopt and apply these values introduces an additional element into the value landscape of science.

Public participation poses a significant threat to the value norms upheld by the mid-twentieth-century SCS and VFI. First, it questions the earlier assumption that unrestrained science is inherently desirable (Weingart, Joubert, and Connaway 2021, 3). In many respects, these participatory projects are inherently constrained and contextually situated, compelling researchers to reflect on the social implications of their work. Second, by granting publics active roles within scientific processes, this model facilitates the integration of public interests into science, dismantling the basic/applied divide. Furthermore, as expectations for scientists evolve from a paradigm of freedom without responsibility to freedom with responsibility (American Association for the Advancement of Science (AAAS) 2017; International Science Council (ISC) 2021), participants can assist scientists, who, as niche experts, often lack the insight or capacity to fully understand which societal values are pertinent when faced with inductive risks. As noted by Koskinen, “When research is scrutinized from diverse value perspectives, it is unlikely that some widely shared values could lead to errors that would remain unnoticed” (2022, 483). Thus, as publics participate more in science, these diverse perspectives can help ensure that scientists account for a broader spectrum of values, ultimately enhancing understanding of responsibilities to society.

38.4 Participant-led science

The public participation model of science communication encompasses a diverse array of practices, yet in all instances, these endeavors are led by scientists. Alternatively, there are forms of science communication—often bordering on being called science itself—where non-formally trained individuals take the reins of research. This approach is sometimes referred to as public-led citizen science or participant-led science. Such initiatives empower publics to formulate research questions, determine and implement methods, analyze data, and draw conclusions from their findings. As an example, take the Open Insulin Foundation (formally the Open Insulin project). This non-profit collaboration led by a team of biohackers, comprising several organizations and community labs across the world aim to develop the first practical, small-scale, and community-centered model for insulin production. Unlike traditional scientific endeavors, participant-led science does not necessarily adhere to the hallmarks of conventional science, such as hypothesis-driven research and institutional systems for quality control (Elliott and Rosenberg 2019). Consequently, these initiatives represent a more open-ended approach to science. They challenge traditional assumptions of epistemic authority and operate outside the conventional norms of scientific practice in terms of expertise, responsibility, and values.

Communication within the participant-led approach is multidirectional and less hierarchical than the deficit model or public participation endeavors. Its alternative strategy fosters an environment

where publics are encouraged to share knowledge with each other, anticipating that others will want to engage with the research. As a result, participant-led initiatives are frequently designed for collaboration and exemplified by open science initiatives. However, unlike conventional science, participant-led research differs in that the participants are not obliged to communicate through established channels (e.g., science journals) to be seen as legitimate by stakeholders. Instead, the research can be self-directed, self-serving, and selectively communicated. For example, an Indigenous community leading a research project can be open to sharing general information, like local fish stock levels, but be secretive about the spawning location of the fish to protect the community's livelihood and heritage. Thus, one potential challenge for participant-led science, given its malleability to context-dependent values, is to generate information that is broadly useful. But since this sort of science communication can be rightfully self-serving, the scope of its value-framing should not be taken as a deterrent to its pursuit, though perhaps a limitation to interest in communicating it.

The increased prominence of participant-led research has elicited both normative and epistemic criticism. Detractors contend that publics may lack the necessary disinterest to yield reliable results, allowing their personal and collective values to skew research toward advocacy-driven conclusions, unduly influenced by ideological conflicts of interest (Resnik, Elliott, and Miller 2015). However, akin to the challenge posed by inductive risk to the VFI, values in science can play a constructive role in science. The values that publics infuse into their projects can highlight the limitations of traditional research choices. For instance, individuals directly impacted by environmental pollution may assess risks of error differently than scientists who are not part of the potentially affected communities (Brown 1992). And, reassuringly, “the philosophical literature suggests that it is not inherently problematic for those engaged in citizen science to bring a perspective of advocacy to their work” (Elliott and Rosenberg 2019, 5) since the values underpinning advocacy need not compromise scientific objectivity (Elliott and Resnik 2014).⁵

When publics decide on research questions, methodologies, thresholds of evidence, and the interpretation of results, they assume additional responsibilities that extend beyond those of mere participants. In traditional scientific practice—and even within the realm of public participation—it is ultimately scientists who bear responsibility for the intended consequences and the foreseeable unintended consequences of their research (Douglas 2003, 2024). However, I posit that when publics adopt these responsibilities, due to their lack of expertise and the inherent limitations of localized projects (Storksdieck et al. 2016), their ability to foresee relevant consequences can be limited. Douglas might respond that publics have no special requirement, beyond what is asked of anyone, to be responsible for the unintended but reasonably foreseeable consequences of their work. But, given these challenges, future discourse should critically examine whether it is reasonable to assign the same level of societal responsibility to publics as is expected of scientists who undergo years of rigorous training, certification, and continued career oversight. If we are to hold the public to such standards, we must also consider how they might acquire the necessary training to effectively foresee the consequences of inquiry and their communication thereof.

A practical response to this call is already taking shape through the emergence of organizations dedicated to educating publics about scientific risks. As this form of research gains traction and matures into a more established practice, both professionals and enthusiasts have united to form membership organizations and associations to reflect upon research risks. These groups strive to create systematic opportunities for networking, codifying practices, and sharing concerns (Stebbins 1992). They also recognize that local contexts significantly influence how participant-led research is valued, implemented, and communicated, fostering socially situated approaches to value-based risk assessment. In this evolving landscape, the interplay between local considerations and broader societal implications invites a critical reexamination of the risks inherent in research, as traditional boundaries of inquiry and responsibility are being reconstructed.

For instance, the US-based Association for Advancing Participatory Sciences, formerly known as the Citizen Science Association, supports individuals who design, lead, manage, and study citizen science initiatives. In Europe, organizations like the European Citizen Science Association seek to identify emerging ethical, social, and legal challenges associated with open science and citizen science. What these organizations have in common is their emphasis on individual investigators and attention to the values investigators bring into multiple stages of the research process—values that, under other science communication models, were assumed to be avoidable or limited to professional scientists.

In sum, participant-led research holds the promise of integrating a diverse array of interests and values into science at multiple stages of the research process. However, this inclusivity brings with it complex challenges, particularly regarding the assignment of responsibility for the social consequences of scientific inquiry. Ongoing discussions surrounding these issues encompass both descriptive and normative questions, necessitating input from philosophers, science communication scholars, and practitioners as well as publics.

38.5 Conclusion

The evolution of science communication reflects shifting ideals about the relationship between science and society. Embedded within these approaches are normative assumptions about scientists, communicators, and publics, which structure how these actors are encouraged to engage with and respond to science. I began by outlining how the mid-twentieth-century SCS and VFI established the deficit model for science communication. The deficit model reinforces the SCS by further isolating basic scientists, who are tasked solely with transmitting information. Such unidirectionality not only underscores the epistemic authority of scientists but also enables them to present their work as value-free and sidestep confrontation about the societal consequences of research.

Criticism of the deficit model paved the way for the emergence of the dialogue model, though the earlier model never disappeared from use entirely. The dialogue model provides the public with its first institutionally sanctioned opportunities to pose questions about science, all while maintaining the authority of science. Furthermore, it replaces the notion of a monolithic public and embraces the idea of diverse publics. Through these inquiries, public interests, attitudes, and values are introduced into scientific discourse. This shift represents a step toward recognizing the importance of public engagement in shaping public perceptions of science. When questioning was found to be insufficient, dialogue evolved into action, and the public participation model emerged. Public participation distinguished itself from the dialogue model by enabling scientists to curate opportunities for public involvement that extended beyond mere questioning. With respect to values, this model accepts social and personal values into science through active public participation in the various stages of the research process. But experimental design and responsibility remain primarily with scientists.

Finally, participant-led research upsets traditional accounts of value exclusion in science by integrating public values throughout the scientific process. However, with publics leading the research process, and public values permeating every stage, this approach raises critical questions about responsibility in science—questions that publics cannot evade. Clarifying how public values can and should influence scientific inquiry, as well as the implications for our understanding of responsibility, has direct implications for the new demarcation problem, new social contracts for science, and the trust placed in scientific endeavors. These approaches to science communication, the management of values within them, and the unresolved questions surrounding the sharing of scientific knowledge should compel philosophers to engage with science communication's modalities to enrich our understanding of science, society, and values.

Notes

- 1 For starting points on science outside the West, see Schirmacher (2013) for insights into Europe and Shiba-rshina and Maslanov (2020) for perspectives on the Soviet Union.
- 2 For articulations, see Levi's (1960) value-neutrality thesis discouraging scientists from considering the societal implications of their work, Price's *The Scientific Estate* (1965), and Wood's "Science and Politics: The Rise of an Apolitical Elite" (1964) for articulations.
- 3 The SCS also suffers criticism around the same time due to challenges to the basic/applied distinction, calls for scientists to take on more responsibility, questions about the funding of science, and the desirability of non-epistemic values in science. For more on this, see Douglas and Branch (2024).
- 4 See Braun and Schultz's (2010) typology of "publics," including the general public, pure public, affected public, and partisan public.
- 5 Of course, publics might not be fully aware or able to articulate the ways in which values influence their research; this is a problem that extends beyond participant-led research and is a greater issue for science itself. (Also see Doucet and Sismondo (2008); Sismondo (2018) on industry-funded research).

References

- Ahteensuu, Marko. 2012. "Assumptions of the Deficit Model Type of Thinking: Ignorance, Attitudes, and Science Communication in the Debate on Genetic Engineering in Agriculture." *Journal of Agricultural and Environmental Ethics* 25 (3): 295–313. <https://doi.org/10.1007/s10806-011-9311-9>.
- American Association for the Advancement of Science (AAAS). 2017. "AAAS Statement on Scientific Freedom and Responsibility." <https://www.aaas.org/programs/scientific-responsibility-human-rights-law/aaas-statement-scientific-freedom> (accessed March 17, 2024).
- Aydeniz, Mehmet, and Sherry A. Southerland. 2012. "A National Survey of Middle and High School Science Teachers' Responses to Standardized Testing: Is Science Being Devalued in Schools?" *Journal of Science Teacher Education* 23 (3): 233–57. <https://doi.org/10.1007/s10972-012-9266-3>.
- Bauer, Martin W. 2008. "Survey Research and the Public Understanding of Science." In *Handbook of Public Communication of Science and Technology*, edited by Massimiano Bucchi and Brian Trench. Routledge. <https://doi.org/10.4324/9780203928240-14>.
- Besley, John C., and Andrea H. Tanner. 2011. "What Science Communication Scholars Think about Training Scientists to Communicate." *Science Communication* 33 (2): 239–63. <https://doi.org/10.1177/1075547010386972>.
- Bodmer, Walter. 1987. "The Public Understanding of Science." *Science and Public Affairs* 2: 69–90.
- Branch, T. Y., and Heather Douglas. 2024. "Rethinking the Conceptual Space for Science in Society after the VFI." *Philosophy of Science* 91 (5): 1118–27. <https://doi.org/10.1017/psa.2023.130>.
- Braun, Kathrin, and Susanne Schultz. 2010. "... A Certain Amount of Engineering Involved": Constructing the Public in Participatory Governance Arrangements." *Public Understanding of Science* 19 (4): 403–19. <https://doi.org/10.1177/0963662509347814>.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-Free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- Brown, Phil. 1992. "Popular Epidemiology and Toxic Waste Contamination: Lay and Professional Ways of Knowing." *Journal of Health and Social Behavior* 33 (3): 267–81. <https://doi.org/10.2307/2137356>.
- Bryant, Chris. 2003. "Does Australia Need a More Effective Policy of Science Communication?" *International Journal for Parasitology* 33: 357–61. [https://doi.org/10.1016/S0020-7519\(03\)00004-3](https://doi.org/10.1016/S0020-7519(03)00004-3).
- Bucchi, Massimiano. 2008. "Of Deficits, Deviations and Dialogues: Theories of Public Communication of Science." In *Handbook of Public Communication of Science and Technology*, edited by Massimiano Bucchi and Brian Trench. Routledge. <https://doi.org/10.4324/9780203928240-11>.
- Burns, Terry W., D. John O'Connor, and Susan M. Stockmayer. 2003. "Science Communication: A Contemporary Definition." *Public Understanding of Science* 12 (2): 183–202. <https://doi.org/10.1177/09636625030122004>.
- Cortassa, Carina. 2016. "In Science Communication, Why Does the Idea of a Public Deficit Always Return? The Eternal Recurrence of the Public Deficit." *Public Understanding of Science* 25 (4): 447–59. <https://doi.org/10.1177/0963662516629745>.
- Davies, Sarah R. 2008. "Constructing Communication: Talking to Scientists about Talking to the Public." *Science Communication* 29 (4): 413–34. <https://doi.org/10.1177/1075547008316222>.
- Doucet, Mathieu, and Sergio Sismondo. 2008. "Evaluating Solutions to Sponsorship Bias." *Journal of Medical Ethics* 34 (8): 627–30. <https://doi.org/10.1136/jme.2007.022467>.

- Douglas, Heather. 2003. "The Moral Responsibilities of Scientists (Tensions between Autonomy and Responsibility)." *American Philosophical Quarterly* 40 (1): 59–68.
- Douglas, Heather. 2024. "Structuring Institutions for Responsible and Accountable Science." *Philosophy of Science* 91 (5): 1149–58. <https://doi.org/10.1017/psa.2023.128>.
- Douglas, Heather, and T. Y. Branch. 2024. "The Social Contract for Science and the Value-Free Ideal." *Synthese* 203: 40. <https://doi.org/10.1007/s11229-023-04477-9>.
- Durant, John R., Geoffrey A. Evans, and Geoffrey P. Thomas. 1989. "The Public Understanding of Science." *Nature* 340 (6228): 11. <https://doi.org/10.1038/340011a0>.
- Elliott, Kevin C., and David B. Resnik. 2014. "Science, Policy, and the Transparency of Values." *Environmental Health Perspectives* 122 (7): 647–50. <https://pmc.ncbi.nlm.nih.gov/articles/PMC4080531/>.
- Elliott, Kevin C., and Jon Rosenberg. 2019. "Philosophical Foundations for Citizen Science." *Citizen Science: Theory and Practice* 4 (1): 1–9. <https://doi.org/10.5334/cstp.155>.
- Finlay, Summer M., Sujatha Raman, Elizabeth Rasekoala, et al. 2021. "From the Margins to the Mainstream: Deconstructing Science Communication as a White, Western Paradigm." *Journal of Science Communication* 20 (1): C02. <https://doi.org/10.22323/2.20010302>.
- Gascoigne, Toss, Donghong Cheng, Michel Claessens, Jenni Metcalfe, Bernard Schiele, and Shunke Shi. 2010. "Is Science Communication Its Own Field?" *Journal of Science Communication* 9 (3): C04. [https://jcom.sissa.it/article/pubid/Jcom0903\(2010\)C04/](https://jcom.sissa.it/article/pubid/Jcom0903(2010)C04/).
- Goldenberg, Maya J. 2016. "Public Misunderstanding of Science? Reframing the Problem of Vaccine Hesitancy." *Perspectives on Science* 24 (5): 552–81. https://doi.org/10.1162/POSC_a_00223.
- Gross, Alan G. 1994. "The Roles of Rhetoric in the Public Understanding of Science." *Public Understanding of Science* 3 (1): 3–23. <https://doi.org/10.1088/0963-6625/3/1/001>.
- Halpern, Megan K., and Kevin C. Elliott. 2022. "Science as Experience: A Deweyan Model of Science Communication." *Perspectives on Science* 30 (4): 621–56. https://doi.org/10.1162/posc_a_00398.
- Hansen, Janus, Lotte Holm, Lynn Frewer, Paul Robinson, and Peter Sandøe. 2003. "Beyond the Knowledge Deficit: Recent Research into Lay and Expert Attitudes to Food Risks." *Appetite* 41 (2): 111–21. <https://www.sciencedirect.com/science/article/abs/pii/S0195666303000795>.
- Holland, Earl M. 2007. "The Risks and Advantages of Framing Science." *Science* 317 (5842): 1168–70. <https://doi.org/10.1126/science.317.5842.1168b>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- Hu, Wanheng. 2024. "Imagining the Model Citizen: A Comparison between Public Understanding of Science, Public Engagement in Science, and Citizen Science." *Public Understanding of Science* 33 (6): 709–24. <https://doi.org/10.1177/09636625241227081>.
- International Science Council (ISC). 2021. "A Contemporary Perspective on the Free and Responsible Practice of Science in the 21st Century: Discussion Paper of the Committee for Freedom and Responsibility in Science."
- Jamieson, Kathleen H. 2017. "The Need for a Science of Science Communication: Communicating Science's Values and Norms." In *The Oxford Handbook of the Science of Science Communication*, edited by Kathleen Hall Jamieson, Dan M. Kahan, and Dietram A. Scheufele. Oxford University Press. <https://academic.oup.com/edited-volume/27956/chapter-abstract/211534625?>
- Koskinen, Inkeri. 2022. "How Institutional Solutions Meant to Increase Diversity in Science Fail." *Synthese* 200 (6): 483. <https://doi.org/10.1007/s11229-022-03959-6>.
- Levi, Issac. 1960. "Must the Scientist Make Value Judgments?" *The Journal of Philosophy* 57 (11): 345–57.
- Miller, Jon D. 1983. "Scientific Literacy: A Conceptual and Empirical Review." *Daedalus* 112: 29–48.
- Moser, Susanne C., and Lisa Dilling. 2011. "Communicating Climate Change: Closing the Science-Action Gap," in *The Oxford Handbook of Climate Change and Society*, edited by John S. Dryzek, Richard B. Norgaard, and David Schlosberg. Oxford University Press. <https://academic.oup.com/edited-volume/28186/chapter-abstract/213097621?redirectedFrom=fulltext&login=false>.
- Nisbet, Matthew C., and Chris Mooney. 2007. "Framing Science." *Science* 316 (5821): 56. <https://www.science.org/doi/10.1126/science.1142030>.
- Price, Don K. 1965. *The Scientific Estate*. The Belknap Press of Harvard University Press.
- Priest, Susanna H. 2001. "Misplaced Faith: Communication Variables as Predictors of Encouragement for Biotechnology Development." *Science Communication* 23 (2): 97–110. <https://doi.org/10.1177/1075547001023002002>.
- Resnik, David B., Kevin C. Elliott, and Aubrey K. Miller. 2015. "A Framework for Addressing Ethical Issues in Citizen Science." *Environmental Science & Policy* 54: 475–81. <https://doi.org/10.1016/j.envsci.2015.05.008>.

- Schirmacher, Arne. 2013. "Introduction: Communicating Science: National Approaches in Twentieth-Century Europe." *Science in Context* 26 (3): 393–404. <https://doi.org/10.1017/S0269889713000112>.
- Shibarshina, Svetlana V., and Evgeby V. Maslanov. 2020. "Science Communication in the Soviet Union: Science as Vocation and Profession." *Social Epistemology* 34 (2): 174–83. <https://doi.org/10.1080/02691728.2019.1695012>.
- Sismondo, Sergio. 2018. *Ghost-Managed Medicine: Big Pharma's Invisible Hands*. Mattering Press. <https://doi.org/10.28938/9780995527775>.
- Stebbins, Robert A. 1992. *Amateurs, Professionals and Serious Leisure*. McGill-Queen's University Press.
- Stocklmayer, Susan. 2013. "Engagement with Science: Models of Science Communication." In *Communication and Engagement with Science and Technology: Issues and Dilemmas—A Reader in Science Communication*, edited by John K. Gilbert and Susan M. Stocklmayer. Routledge.
- Storksdieck, Martin, Jennifer Lynn Shirk, Jessica L. Cappadonna, et al. 2016. "Associations for Citizen Science: Regional Knowledge, Global Collaboration." *Citizen Science: Theory and Practice* 1 (2): 1–10. <https://doi.org/10.5334/cstp.55>.
- Strasser, Bruno J., Jérôme Baudry, Dand Mahr, Gabriela Sanchez, and Elise Tancoigne. 2019. "Citizen Science? Rethinking Science and Public Participation." *Science & Technology Studies* 32 (2): 52–76. <https://doi.org/10.23987/sts.60425>.
- Sturgis, Patrick, and Nick Allum. 2004. "Science in Society: Re-Evaluating the Deficit Model of Public Attitudes." *Public Understanding of Science* 13 (1): 55–74. <https://doi.org/10.1177/0963662504042690>.
- Suldovsky, Brianne. 2016. "In Science Communication, Why Does the Idea of the Public Deficit Always Return? Exploring Key Influences." *Public Understanding of Science* 25 (4): 415–26. <https://doi.org/10.1177/0963662516629750>.
- Trench, Brian. 2008. "Towards an Analytical Framework of Science Communication Models." In *Communicating Science in Social Contexts: New Models, New Practices*, edited by Donghong Cheng, Michel Claessens, Toss Gascoigne, Jenni Metcalfe, Bernard Schiele, and Shunke Shi. Springer. https://doi.org/10.1007/978-1-4020-8598-7_7.
- Trench, Brian, and Kirk, Junker. 2001. "How Scientists View Their Public Communication." In *Sixth International Conference on Public Communication of Science and Technology: Trends in Science Communication Today: Bridging the Gap between Theory and Practice*. Proceedings of the PCST 2001, 1–3.
- van der Sanden, Maarten C. A., and Frans J. Meijman. 2008. "Dialogue Guides Awareness and Understanding of Science: An Essay on Different Goals of Dialogue Leading to Different Science Communication Approaches." *Public Understanding of Science* 17 (1): 89–103. <https://doi.org/10.1177/0963662506067376>.
- Weingart, Peter, Marina Joubert, and Karien Connoway. 2021. "Public Engagement with Science—Origins, Motives and Impact in Academic Literature and Science Policy." *PLoS One* 16 (7): e0254201. <https://doi.org/10.1371/journal.pone.0254201>.
- Wood, Robert C. 1964. "Scientists and Politics: The Rise of the Apolitical Elite." In *Scientists and National Policymaking*, edited by Robert Gilpin and Christopher Wright. Columbia University Press.
- Wynne, Brian. 1992. "Public Understanding of Science Research: New Horizons or Hall of Mirrors?" *Public Understanding of Science* 1 (1): 37–43. <https://doi.org/10.1088/0963-6625/1/1/008>.

VALUES IN RESEARCH ETHICS EDUCATION

Kristina Rolin

39.1 Introduction

This chapter argues that research ethics education would be improved by integrating perspectives from the philosophical literature on values and science. Research ethics studies conceptions of good scientific practice from a normative perspective. Much of research ethics focuses on norms that tell researchers what they ought to do and what they are not permitted to do. A normative inquiry asks which norms should guide researchers' actions in various stages of research and other activities of scientific communities. It also examines the ethical, epistemic, and social justification of norms for conduct. Such an inquiry aims to understand from where research ethical norms derive their normative force, why researchers should follow these norms, and why scientific institutions are justified in fostering, supporting, and enforcing these norms. While the terms “research integrity” and “responsible conduct of research” are often used in connection with research ethics education, this chapter uses the term “research ethics” to invite a broader perspective on the field.

The philosophical literature on values and science aims to understand, among other things, how epistemic values differ from non-epistemic ones, including moral and social values. It is widely believed that epistemic values are at the heart of good scientific practice. It is difficult to understand what scientific practices are about unless they are seen as practices aiming at epistemic goals, including significant truth (Kitcher 2001), empirical success (Solomon 2001), problem-solving (Laudan 1977), and understanding (Potochnik 2020). By epistemic values are meant values that promote the epistemic goals of science. For example, Steel (2010) defines epistemic values as values that promote the attainment of truth, either intrinsically or extrinsically. An epistemic value is intrinsic when manifesting that value constitutes an attainment of or is necessary for truth, and it is extrinsic when it promotes the attainment of truth contingently, without itself being an indicator or a requirement of truth (2010, 15).

While epistemic values are constitutive of good scientific practice, it is less clear what the proper role of moral and social values is in scientific inquiry. By moral values are meant such values as respect for human rights (including respect for the privacy, human dignity and autonomy of research participants), concern for the welfare of laboratory animals, and due care for safety and the environment. By social values are meant such values as social justice, equality, democracy, and sustainable development. Research ethics education assumes that moral and social values should guide and constrain scientific inquiry. Not only should the epistemic ends of scientific inquiry be morally

and socially acceptable but also the means that are used to pursue the ends. However, as this chapter argues, research ethics education would benefit from further clarification on how moral and social values should guide and constrain scientific inquiry.

The philosophical literature on values and science provides such clarification by helping researchers understand what roles moral and social values can and should play in the core activities of scientific inquiry. Core activities mean practices whereby knowledge claims are evaluated and either rejected or accepted and communicated to others. The literature is concerned not merely with norms of conduct but also with ideals for which researchers should strive even when the ideals may be unattainable. Much of the values and science literature argues against the value-free ideal of science, the view that non-epistemic values are not permitted to play a role in the core activities of scientific inquiry. While many philosophers reject the value-free ideal (e.g., [Brown 2024](#); [de Melo-Martín and Intemann 2016](#); [Douglas 2009](#); [Elliott 2017](#); [Longino 1990, 1995](#)), some others attempt to defend the legacy of the value-free ideal by introducing revised versions of the ideal (e.g., [Menon and Stegenga 2023](#)). The literature is also concerned with the question of how the role of moral and social values in science should be managed if the core activities of scientific inquiry cannot be free from such values. The so-called value-management question ([Brown 2024](#)), which is sometimes called “the new demarcation problem” ([Holman and Wilholt 2022](#)), asks when and how moral and social values are permitted to enter scientific judgment and which values are permitted to do so.

This chapter is organized in the following way. [Section 39.2](#) argues that the literature on values and science enhances research ethics education by helping researchers understand the epistemic, moral, and social justification of norms. [Section 39.3](#) argues that the literature clarifies the proper role of moral and social values in the core activities of scientific inquiry. [Section 39.4](#) argues that the literature calls attention to the strengths and weaknesses of institutional policies to increase awareness of the proper role of values in science.

39.2 The epistemic, moral, and social justification of research ethical norms

Research ethics education aims not only to inform researchers about research integrity guidelines and norms; it also helps them understand why such guidelines and norms should be followed. In many countries, national research integrity offices, higher education institutions, research funding agencies, and scientific journals enforce research ethical norms with sanctions on those scientists who violate codes of conduct. For example, research funding agencies can refuse to process applications from scientists found guilty of research misconduct, scientific journals can retract their articles, and universities can fire scientists who have committed serious fraud. As these consequences may function as incentives for researchers to comply with guidelines and norms, they provide one answer to the question of why these guidelines and norms should be followed. The answer is that in many cases, compliance is in the researchers’ self-interest. They should follow research ethical norms to protect their professional interests and reputations. Norm violations can damage their career prospects and result in a waste of resources and time they have invested in their work.

This section seeks more fundamental answers to the question of why research ethical norms should be followed by examining their justification. These answers help mitigate a common concern about research ethics education. The worry is that research ethics appears to be a set of externally imposed rules rather than something intrinsic to scientific practice ([Pennock and O’Rourke 2017](#)). If researchers understand that many norms can be justified on epistemic grounds in addition to moral and social grounds, it is easier for them to see that research ethics is intrinsic to scientific practice.

One way to inquire about the justification of research ethical norms is to ask what goals such norms are meant to serve and why these goals are valuable, either intrinsically or instrumentally to some other goal that is valuable in and of itself. As the very term research *ethics* suggests, many research

ethical norms are meant to promote widely shared moral and social values (e.g., social responsibility, human rights, animal welfare, compliance with the law, and public health and safety). Resnik (2020) argues that in addition to morally and socially valuable goals, compliance with research ethical norms promotes three other goals: (i) The epistemic aims of research (e.g., truth and the avoidance of error); (ii) values that are essential to collaborative work (e.g., trust, accountability, and mutual respect); and (iii) accountability to the public (e.g., disclosing conflicts of interest). Taking my lead from Resnik, I discuss examples of research ethical norms that promote these three goals.

39.2.1 Epistemic aims of research

The prohibition against fabricating, falsifying, or misrepresenting data can be justified epistemically by arguing that it promotes the epistemic goals of science. As research data play a key role in the epistemic justification of scientific knowledge, it is hard to see how fabricated or falsified data could contribute to the reliability of scientific knowledge. The prohibition against fabricating, falsifying, or misrepresenting data can also be justified on moral and social grounds. Fabrication, falsification, and misrepresentation of data are wrong because they are forms of dishonesty, and they result in unfairness in the distribution of credit. According to Resnik, a fair distribution of credit in science amounts to the principle that “credit should be given where credit is due but not where it is not due” (1996, 575). This principle is important in scientific publication practices because publications are the main vehicles of credit in science. While the greatest credit belongs to those researchers who are properly listed as authors of a publication, credit is also given to those whose publications are cited or who are recognized in acknowledgements. Fabrication, falsification, and misrepresentation of data undermine fairness because they enable scientists to accumulate credit they do not deserve. It is much easier to be a prolific researcher by analyzing fake data than by producing authentic data by means of experiments, observations, or other data-gathering methods. While fairness is a social value, it is closely connected to the epistemic goals of science. The scientific community can hardly achieve its epistemic aims unless it maintains a reward structure of science that is seen as fair.

39.2.2 Scientific collaboration

Insofar as research ethical norms promote collaboration, they acquire epistemic justification in virtue of serving the epistemic aims of science. In many areas of science, collaboration is a practical necessity because the production and analysis of data are too demanding, expensive, or time-consuming for any individual scientist to accomplish independently (Hardwig 1991). Sometimes collaboration is a necessity because a research project requires expertise and skills from different disciplines (Andersen and Wagenknecht 2013). In such cases, a research team with a division of labor can carry out a project that no individual scientist could do on their own.

As collaboration often results in co-authored publications, research ethics must consider the question of who is entitled to authorship. The philosophical literature on values and science can improve research ethics education by clarifying the moral and epistemic justification for authorship policies. A moral justification for authorship policies is based on the above-mentioned principle for a fair distribution of credit. Unfair denial of authorship (someone who has made a substantial contribution is not offered an opportunity to become a co-author), guest authorship (someone who does not qualify as author is listed as such), and ghost writing (someone has agreed to write on someone else’s behalf as an unnamed author) are morally wrong because they undermine a fair distribution of credit in addition to being forms of dishonesty. The fairness principle is violated because those who deserve credit do not receive it, and some others receive more than their fair share of credit. Ghost writing is also morally problematic because it is a way to hide a conflict of interest that should be disclosed.

While the moral rationale for authorship policies may seem to be obvious, the epistemic rationale is not equally easy to understand. The epistemic rationale is based on an obligation to take epistemic responsibility for one's knowledge claims. Epistemic responsibility requires that one provides evidence or other arguments in support of one's claims, or at least that one commits oneself to defending the claims, for example, by responding to challenges and further inquiries (Rolin 2017). An obligation to be epistemically responsible arises when one agrees to be a co-author of a scientific publication. Violations of authorship policies are likely to lead to epistemically irresponsible behavior. Unfair denial of authorship, honorary authorship, ghost writing, and plagiarism give rise to a situation where not all listed authors can carry out their epistemic responsibilities. These practices are epistemically problematic because they make it impossible for others to locate those persons who could take epistemic responsibility for scientific claims. An obligation to take epistemic responsibility for one's knowledge claims is also implicit in the ICMJE (International Committee of Medical Journal Editors) guidelines, emphasizing that authorship involves an agreement to "be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved" (International Committee of Medical Journal Editors (ICMJE) 2024).

In addition to authorship policies, the moral values of honesty, sincerity, and benevolence play an epistemic role in science in virtue of facilitating scientific collaboration. They do so by helping scientists build and maintain mutual relations of trust. Many philosophers argue that mutual relations of trust are necessary for successful collaboration (Frost-Arnold 2013). Scientists would hardly collaborate if they did not trust their team members to do their share in the collective effort. Moreover, scientists often trust their collaborators as a source of knowledge. Thus, mutual relations of trust are necessary both for practical and epistemic reasons. To foster mutual relations of epistemic trust, collaborating scientists must strive to be epistemically trustworthy. In addition to having a relatively high degree of domain-specific expertise, an epistemically trustworthy person is honest, sincere, and has goodwill toward those who are epistemically dependent on her (Rolin 2024a). While honesty, sincerity, and benevolence are moral virtues, they are also extrinsic epistemic values because they promote collaboration among scientists (Rolin 2015).

Scientific collaboration in a broad sense involves cooperative activities that are necessary to create and maintain epistemically well-functioning scientific communities. By advising scientists to respect their colleagues, research ethics guidelines aim to promote such communities. The philosophical literature on values and science can enhance this aspect of research ethics education in significant ways. For example, Longino argues that for scientific communities to function well epistemically, they must uphold the norms of public venues, uptake of criticism, shared standards, and equality of intellectual authority (1990, 76–80). According to Longino, the exclusion of women and members of minorities from scientific education and the scientific profession constitutes a violation of the fourth norm (1990, 78). While the fourth norm is in line with important moral and social values (e.g., the principle of non-discrimination), it is also justified on epistemic grounds. Longino argues that the fourth norm as well as the three others promote critical interactions among scientists, thereby helping them identify and correct false beliefs or biased research (1990, 80).

39.2.3 Accountability to the public

Accountability to the public is important because scientific education and research are largely funded publicly, and research depends on the willingness of citizens to participate in it. Like some other social values (e.g., fairness), accountability to the public is connected to the epistemic goals of science. The scientific community can hardly continue to pursue its epistemic aims if it fails to be accountable to the public.

A requirement to disclose conflicts of interest is an example of a norm meant to promote accountability to the public. A conflict of interest arises when someone has a financial, personal, political, or other interest that may bias their judgment or decision-making concerning the performance of their obligations as a researcher. As it is often impossible to avoid conflicts of interest, research ethics guidelines require that they be disclosed to potential research participants, the scientific community, and citizens interested in research. Having and declaring a conflict of interest is not seen as a research ethical wrongdoing, but hiding one amounts to disregard for good scientific practice.

The literature on values and science suggests that there are limitations to what conflicts-of-interest policies can achieve. It is not clear whether disclosing conflicts of interest makes researchers more attentive to possible bias resulting from their interests and values (de Melo-Martín 2023). As Longino argues, one challenge is to identify values that are invisible to most members of scientific communities (1990, 80). Moreover, it is not clear what is achieved by shifting the burden of identifying bias to citizens interested in research. If scientists have a hard time detecting bias in their research, lay people are hardly better equipped to do so. While there is evidence suggesting that conflicts of interest increase the risk of bias (Elliott 2014), lay people may not understand how research is biased. The philosophical literature on values and science reveals that research ethics education has had too much faith in the ability of conflict-of-interest policies to increase the objectivity of scientific knowledge. To promote objectivity, research ethics education should provide researchers with analytic tools that help them identify often tacit value judgments and examine them critically.

While the philosophical literature on values and science calls attention to the limitations of conflict-of-interest policies, it can contribute to research ethics education in a positive way by inviting a broader approach to accountability. A broader approach goes beyond accountability as compliance with legal and ethical norms to include socially responsible science (de Melo-Martín and Intemann 2023; Douglas 2024). While research ethics guidelines inform scientists about minimum standards they need to maintain, social responsibility recommends them to do more than to meet minimum standards. For example, Kourany defends the ideal of socially responsible science, which means that “sound social values, as well as sound epistemic values, must control every aspect of the scientific research process, from the choice of research questions to the communication and application of results” (2010, viii). Given Kourany’s ideal, the challenge is to understand not only what appropriate social values are but also how scientists are informed about these values. In pluralist societies, there can be reasonable disagreements over social values (as long as the values do not contradict the core principles of liberal democracy), and it is not clear whose social values scientists should serve (Rolin 2021). In response to this challenge, some philosophers propose procedural accounts of social responsibility aiming to ensure that scientists’ social value judgments conform to the democratic decisions of ideal deliberators (Kitcher 2001) or actual deliberators (Lusk 2021).

In sum, the philosophical literature on values and science improves research ethics education by showing that many research ethical norms are hybrid norms in the sense that they serve both epistemically and morally or socially valuable goals. While research ethics is sometimes seen merely as a branch of applied ethics, it is also part of the social epistemology of science.

39.3 The proper role of moral and social values in the core activities of scientific inquiry

Research ethics education would benefit from clarification on the proper role of non-epistemic values in scientific inquiry. As Ambrosj, Desmond, and Dierickx argue, one weakness of research integrity guidelines is “a lack of clear guidance for researchers on how to integrate non-epistemic values in their research” (2023, 17). They have conducted a content analysis of national research integrity

guidelines in 32 European countries, focusing on guidance on non-epistemic values in research. They argue that research integrity guidelines divide non-epistemic values roughly into three categories. First, there are non-epistemic values that can legitimately influence scientific inquiry. This category includes honesty, care, respect, transparency, openness, confidentiality, cautiousness, accountability, collegiality, fairness, and justice. Second, there are non-epistemic value influences that are illegitimate and, therefore, should be avoided or minimized. In this category, they list external influences, conflicts of interest, economic interests, self-interests, politics, pressure to publish, and personal beliefs. Third, there are non-epistemic values that are not illegitimate as such but potentially problematic, and therefore, should be disclosed. The third category overlaps in part with the second category and includes conflicts of interest, economic interests, self-interests, and personal beliefs. Ambrosj, Desmond, and Dierickx conclude that research integrity guidelines offer mixed messages about appropriate non-epistemic values in research.

To make things even more confusing, research integrity guidelines endorse positions that come close to the value-free ideal of science, “passages claiming that researchers should be free and independent from any external influence” (Ambrosj, Desmond, and Dierickx 2023, 1). A prominent example of such a position can be found in the *European Code of Conduct for Research Integrity*, which states that “research draws on the work of the community of researchers and should develop independently of pressure from commissioning parties and from ideological, economic, or political interests” (All European Academies (ALLEA) 2023, 3). The passage raises the question of whether research can proceed fully independently of such pressures and interests, and if it cannot, how such values influence in science should be managed.

The philosophical literature on values and science has a lot to say in response to this question. This section discusses three influential arguments against the value-free ideal of science: (i) An argument from value-laden background assumptions, (ii) an argument from plurality of epistemic values, and (iii) an argument from inductive risk. The three arguments aim to show that in many cases the value-free ideal is not feasible, and even when it is feasible, it is not desirable because it runs contrary to important moral principles. The values and science literature improve research ethics education because it shows that when researchers cannot avoid making moral and social value judgments, they should make such judgments explicitly and in a responsible way. As Elliott argues, value influences should be made as transparent as possible, they should be representative of major social and ethical priorities, and they should be scrutinized through appropriate processes of engagement between different scholars and stakeholders (2017, 12).

39.3.1 Argument from value-laden background assumptions

Longino (1990) argues that the value-free ideal is difficult to attain because non-epistemic values can legitimately enter scientific inquiry via background assumptions. In her view, the acceptance of scientific knowledge claims takes place in a context of background assumptions, which are necessary to establish the relevance of empirical evidence to a hypothesis or a theory (1990, 43–44). In some cases, non-epistemic values are embedded in these assumptions. Longino argues further that the value-free ideal is not desirable because non-epistemic values can play epistemically positive roles in science. When different values are present in scientific communities, they are likely to generate different perspectives on the subject matter of inquiry, and such a diversity is epistemically beneficial because it promotes critical interactions among scientists (1990, 80). Scientists are more likely to identify and criticize value-laden background assumptions in research when the values in question are different from their own. Longino’s argument against the value-free ideal of science improves research ethics education by calling attention to moral and social value judgments embedded in theoretical frameworks and concepts.

39.3.2 *Argument from plurality of epistemic values*

Longino (1995) also argues that the value-free ideal is difficult to attain because non-epistemic values can legitimately influence decisions to emphasize certain epistemic values in the evaluation of theories. Epistemic values are a diverse set of criteria and desiderata, including theoretical virtues such as empirical adequacy, internal and external consistency, and explanatory power. There may be trade-offs between epistemic values (e.g., between empirical adequacy and simplicity), such that scientists cannot maximize two values at the same time. Longino's argument improves research ethics education by urging researchers to reflect on moral and social value influences in their decisions to consider some epistemic values as weightier than others.

An argument from plurality of epistemic values also calls attention to the ethics of scientific modeling, which is an understudied topic in research ethics (Rolin 2024b). The ethics of scientific modeling is concerned, for example, with the question of how idealization can be justified, so that it does not amount to falsification and research misconduct. As many philosophers argue, the construction of models necessarily involves idealization (Potochnik 2020). Idealization promotes understanding because phenomena studied in science are complex, whereas human cognitive capacities are limited. Idealized models may even include assumptions that scientists know to be false, thereby challenging the received view that scientific inquiry aims merely at truth (Potochnik 2020, 940). Parker has developed this insight further by arguing that scientific models should not be assessed by focusing merely on their representational accuracy but by asking whether they are "adequate for a purpose" (2020, 467). As models can be constructed for many purposes, it is hard to evaluate their adequacy independently of non-epistemic values (Rolin 2024b, 383). The ethics of scientific modeling is concerned with the legitimacy of these non-epistemic values.

39.3.3 *Argument from inductive risk*

Douglas (2009) argues that even if the value-free ideal is feasible, it is not desirable because scientists have a moral obligation to assess risks involved in either rejecting or accepting scientific knowledge claims. Accepting knowledge claims involves uncertainty, and scientists must decide when the evidence at hand is sufficiently strong to warrant acceptance. Reporting knowledge claims with uncertainties, thereby refusing to accept or reject anything, is not an option when scientists are expected to offer informative and relevant knowledge. Whether a knowledge claim should be accepted and communicated to others depends on the consequences. If a scientist accepts a false hypothesis, there may be harm associated with this type of error. If she rejects a true one, there may be another harm associated with the other type of error. The assessment of the seriousness of these harms is a moral and social value judgment. The value-free ideal is unacceptable because it would encourage scientists to ignore their moral obligation to avoid causing harm or at least to minimize it.

Douglas's (2009) argument enhances research ethics education significantly by providing a more refined analysis of risks than what can be found in the existing research ethics literature. The existing literature emphasizes that researchers must recognize those persons and stakeholder groups who might be impacted by research (e.g., human participants in research, collaborators, and the users of knowledge). They must also identify foreseeable consequences and assess whether they promote well-being and a fair distribution of welfare among those who are affected by research. Most importantly, researchers must envision potential harms and risks and take measures to avoid harm or to minimize risks and mitigate them if they are realized. An argument from inductive risk shows that researchers must also consider the foreseeable consequences of errors they might have made in research. While it is difficult to foresee all the purposes for which research knowledge might be used in the future, it is reasonable to expect that researchers can consider some scenarios where research

knowledge is used in political decision-making or in the development of medicine, treatments, and novel technologies.

Before concluding, it is necessary to comment on an objection that is often raised against Douglas's argument. The objection is that by making moral and social value judgments, scientists assume a role they are not entitled to in democratic societies. Although [de Melo-Martín and Intemann \(2016\)](#) do not endorse the value-free ideal of science, they believe that the critics of the value-free ideal must respond to concerns about the proper role of scientific experts in democratic societies. As they explain the concern, "Scientists have no special expertise or authority in making ethical and social value judgments, and thus, it is unclear that they alone should be the ones deciding what values ought to be endorsed" (2016, 503). In response to this concern, I argue that the worry about scientists' moral and social value judgments threatening democratic principles is exaggerated. Research ethics education can inform scientists about widely shared moral and social values that do not contradict democratic principles. Scientists do not assume the role of a policymaker by applying uncontroversial moral and social values to their research; instead, they aim to fulfill their moral obligations. Moreover, the critics of the value-free ideal of science have proposed various procedures scientists can follow to align their social value judgments with the values of those who are impacted by the application of research knowledge ([Lusk 2021](#)). The critics have also argued that science would benefit from increased social diversity, which is one way to introduce diverse value perspectives in scientific communities ([Longino 1990](#)).

In sum, the philosophical literature on values and science improves research ethics education by showing that moral and social value judgments can influence theoretical, methodological, and conceptual choices. There are no epistemic reasons to ban moral and social values from the core activities of scientific inquiry. While scientists' values can give rise to biases, they can also help them detect biases in other scientists' works. Non-epistemic values are not epistemically bad or good (or neutral) in and by themselves. They can drive scientific inquiry into different directions by either increasing or decreasing biases. There are no moral reasons to ban moral and social values from core activities either. Researchers have a moral obligation to assess the foreseeable consequences of their actions, including the consequences of error.

39.4 Institutional policies to increase awareness of values in science

Research integrity guidelines address not only individual researchers but also scientific institutions and propose various policies they can implement to promote compliance with research ethical norms. This section examines the question of whether these policies can also be used to promote awareness of the proper role of values in science. There are four things that scientific institutions can do to influence their members' behavior. They can (i) make a public commitment to uphold appropriate values, (ii) allocate resources for education and advising, (iii) implement sanctions for those researchers who violate norms, and (iv) require that research plans and protocols be submitted to an ethical review. I argue that the first two policies are promising avenues for improving understanding of the proper role of values in science, whereas the other two are not.

39.4.1 Public commitment

To influence their members' behavior, scientific institutions can communicate their values explicitly in public websites, official publications, and speeches at public events in the community. The values and science literature could improve institutional value statements by providing justification for research ethical norms in addition to merely listing values, which seems to be a common practice. However, there are limitations to what these communications can achieve, even if they are improved.

Unless scientific institutions implement other policies, their value statements run the risk of appearing as a mere lip-service commitment to important epistemic, moral, and social values.

39.4.2 Education and advising

Higher education institutions can promote understanding of values and science by organizing research ethics education and nominating advisors who can provide confidential research ethical advice to researchers. The literature on values and science can enhance the content of research ethics education and advising in many ways. However, education also faces practical challenges. For example, institutions need to decide whether education is voluntary or mandatory. Voluntary education runs the risk that it reaches only those scientists who already understand the importance of values in science. Mandatory education runs the risk of being resisted by those scientists who see it as an extra burden in their workload. The challenge is to ensure that education on values and science is not just a box-ticking exercise. While education is necessary to promote understanding of values and science, it is only one step toward an organizational culture that encourages reflection on values and science.

39.4.3 Sanctions for norm violations

To implement sanctions for those scientists who violate research ethical norms, scientific institutions must have clear definitions of research misconduct and disregard for good scientific practice. They also must define procedures for investigating alleged violations, and make sure that the rights of all parties are safeguarded by a fair process and that all parties have a chance to be heard. While sanctions are justified in cases of research misconduct and disregard for good scientific practice, they are not an appropriate way to promote awareness of the proper role of values in scientific inquiry. The reason for this is that it is difficult to defend precise rules for the proper use of values in science. As Douglas argues, imprecise rules could open science to “serious risks of harmful politicization” (2024, 3). Sanctions, she argues, belong to “accountability mechanisms,” and such mechanisms are inapt for other purposes (8).

39.4.4 Ethical review

Scientific institutions can require some types of research to be reviewed by an ethics committee (or an institutional review board). Typically, such a review must be done before researchers start gathering data. In the case of medical research with human participants and research with laboratory animals, an ethical review is often a legal requirement. In the case of other research with human participants (e.g., in education, psychology, and social sciences), it is part of the self-regulation of the scientific community. In such cases, a failure to acquire a positive ethical review statement may prevent researchers from publishing their results in high-quality scientific journals. If mandatory ethical review were used as a tool to increase awareness of values and science, it would have to be tailored for this purpose. The guidelines for ethical review could ask researchers to reflect on the role of moral and social values in research more broadly than what current research ethical guidelines do.

However, while the requirement for ethical review is justified in cases of research with human participants or with laboratory animals, it is not an ideal way to promote awareness of the literature on values and science. It is not clear that members of ethics committees can make a better analysis of the role of moral and social value judgments in research than scientists themselves. If committee members are not informed about the proper role of values in science, an ethical review can even make things worse by giving positive ethical statements to projects that are problematic from a moral and social point of view. And even if members of ethics committees actively educate themselves about

the proper role of values in science, the implementation of ethical review is a significant cost for scientific institutions as well as for scientists who are required to apply for a review. Scientific institutions would have to consider carefully whether the benefits outweigh the costs.

In sum, institutions can foster and support awareness of the proper role of values in science by means of public commitments, education, and advising, but attempts to enforce such awareness by means of sanctions or a mandatory ethical review may not be successful.

References

- ALLEA (All European Academies). 2023. *The European Code of Conduct for Research Integrity: Revised Edition 2023*. <https://allea.org/wp-content/uploads/2023/06/European-Code-of-Conduct-Revised-Edition-2023.pdf>.
- Ambrosj, Jacopo, Hugh Desmond, and Kris Dierickx. 2023. "The Value-Free Ideal in Codes of Conduct for Research Integrity." *Synthese* 202: 133. <https://doi.org/10.1007/s11229-023-04377-y>.
- Andersen, Hanne, and Susann Wagenknecht. 2013. "Epistemic Dependence in Interdisciplinary Groups." *Synthese* 190: 1881–98. <https://doi.org/10.1007/s11229-012-0172-1>.
- Brown, Matthew J. 2024. "For Values in Science: Assessing Recent Arguments for the Ideal of Value-free Science." *Synthese* 204: 112. <https://doi.org/10.1007/s11229-024-04762-1>.
- de Melo-Martin, Inmaculada. 2023. "Disclosing Conflicts of Interest to Potential Research Participants: Good for Nothing?" *Ethics & Human Research* 45 (2): 2–13. <https://doi.org/10.1002/eahr.500157>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2016. "The Risk of Using Inductive Risk to Challenge the Value-Free Ideal." *Philosophy of Science* 83 (4): 500–20. <https://doi.org/10.1086/687259>.
- de Melo-Martin, Inmaculada, and Kristen Intemann. 2023. "Socially Responsible Science: Exploring Complexities." *European Journal for Philosophy of Science* 13: 33. <https://doi.org/10.1007/s13194-023-00537-6>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Douglas, Heather. 2024. "Structuring Institutions for Responsible and Accountable Research." *Philosophy of Science* 91 (5): 1149–58. <https://doi.org/10.1017/psa.2023.128>.
- Elliott, Kevin C. 2014. "Financial Conflicts of Interest and Criteria for Research Credibility." *Erkenntnis* 79 (Suppl 5): 917–37. <https://doi.org/10.1007/s10670-013-9536-2>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Frost-Arnold, Karen. 2013. "Moral Trust & Scientific Collaboration." *Studies in History and Philosophy of Science Part A* 44 (3): 301–10. <https://doi.org/10.1016/j.shpsa.2013.04.002>.
- Hardwig, John. 1991. "The Role of Trust in Knowledge." *Journal of Philosophy* 88 (12): 693–708. <https://www.jstor.org/stable/2027007>.
- Holman, Bennett, and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- ICMJE (International Committee of Medical Journal Editors). 2024. "Defining the Role of Authors and Contributors." <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford University Press.
- Kourany, Janet. 2010. *Philosophy of Science after Feminism*. Oxford University Press.
- Laudan, Larry. 1977. *Progress and Its Problems: Toward a Theory of Scientific Growth*. Routledge and Kegan Paul.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. *Science as Social Knowledge*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Longino, Helen E. 1995. "Gender, Politics and Theoretical Virtues." *Synthese* 104 (3): 383–97. <https://doi.org/10.1007/BF01064506>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90: 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Menon, Tarun, and Jacob Stegenga. 2023. "Sisyphian Science: Why Value Freedom is Worth Pursuing." *European Journal for Philosophy of Science* 13: 48. <https://doi.org/10.1007/s13194-023-00552-7>.
- Parker, Wendy S. 2020. "Model Evaluation: An Adequacy-for-Purpose View." *Philosophy of Science* 87 (3): 457–77. <https://doi.org/10.1086/708691>.

- Pennock, Robert T., and Michael O'Rourke. 2017. "Developing a Scientific Virtue-based Approach to Science Ethics Training." *Science and Engineering Ethics* 23: 243–62. <https://doi.org/10.1007/s11948-016-9757-2>.
- Potochnik, Angela. 2020. "Idealization and Many Aims." *Philosophy of Science* 87 (5): 933–43. <https://doi.org/10.1086/710622>.
- Resnik, David. 1996. "Social Epistemology and the Ethics of Research." *Studies in History and Philosophy of Science Part A* 27 (4): 565–86. [https://doi.org/10.1016/0039-3681\(96\)00043-X](https://doi.org/10.1016/0039-3681(96)00043-X).
- Resnik, David. 2020. "What Is Ethics in Research and Why Is It Important?" Updated December 23, 2024. <https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm>.
- Rolin, Kristina. 2015. "Values in Science: The Case of Scientific Collaboration." *Philosophy of Science* 82 (2): 157–77. <https://doi.org/10.1086/680522>.
- Rolin, Kristina. 2017. "Scientific Community: A Moral Dimension." *Social Epistemology* 31 (5): 468–83. <https://doi.org/10.1080/02691728.2017.1346722>.
- Rolin, Kristina. 2021. "Objectivity, Trust and Social Responsibility." *Synthese* 199: 513–33. <https://doi.org/10.1007/s11229-020-02669-1>.
- Rolin, Kristina. 2024a. "Earning Epistemic Trustworthiness: An Impact Assessment Model." *Synthese* 203: 39. <https://doi.org/10.1007/s11229-023-04472-0>.
- Rolin, Kristina. 2024b. "Models and Values." In *The Routledge Handbook of Philosophy of Scientific Modeling*, edited by Tarja Knuuttila, Natalia Carrillo, and Rami Koskinen. Routledge. <https://www.taylorfrancis.com/chapters/oa-edit/10.4324/9781003205647-33/models-values-kristina-rolin>.
- Solomon, Miriam. 2001. *Social Empiricism*. The MIT Press.
- Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77 (1): 14–34. <https://doi.org/10.1086/650206>.

PART VI

Values in Specific Sciences



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

“EVERY CELL HAS A SEX”

Sex Essentialism in Biomedical Research

Robyn Bluhm

This chapter brings together two areas of feminist concern to illuminate the values underlying the US NIH’s policy on “sex as a biological variable” (SABV). Helen Longino (1990) identifies three “patriarchal values” that might influence research on sex differences: Sexism, androcentrism, and sexual essentialism. I will trace the influence of these values in sex difference research. Ultimately, I argue, recent efforts to overcome androcentrism in biomedical research have reinforced the field’s existing sex essentialist views.

The first problem about which feminists (including bioethicists) have raised concerns is related to women’s health. There is clear evidence of a variety of health disparities between women and men, and despite these epidemiological facts being well-documented, neither the reasons for these differences nor their implications for prevention or treatment are well-understood. However, these health disparities have been traced, in part, to an underrepresentation of women as participants in biomedical research. More recently, advocates have turned their attention to preclinical research and have successfully lobbied for policy changes designed to ensure that both female and male animals are included in preclinical research and that potential sex differences are explicitly studied. This expansion, however, brings biomedical research into tension with a second area of feminist concern. Feminist critics of sex difference research (including philosophers of science) have pointed out the many ways that scientists have relied on stereotypes about women, and their differences from men, in their empirical research and in their theories. Often, these stereotypes are rooted in value judgments about women, men, and the appropriate relationship between them.

In the next two sections, I will expand on these two views on sex difference research, starting with efforts to increase women’s participation in clinical research, and the expansion of these efforts to preclinical research. I then offer a different perspective on sex difference research by tracing its sexist history and feminist analyses that suggest that it continues to be informed by gender stereotypes. In doing so, I introduce the framework that guides research on sex differences, the organizational-activational hypothesis, and then show how this framework, and its expansion over the years, have reinforced sex essentialism.

40.1 Battling androcentrism: Advocating for research on sex differences

Philosophers, medical researchers, and health policymakers all agree that research on women’s health lags behind research on men’s health, despite changes in policy and practice over the past few

decades. For example, the Society for Women's Health Research identifies a number of health issues specific to women, as well as health concerns more prevalent in women, that contribute to health disparities.¹ Health issues specific to women, such as those related to menopause, are under-researched. Diseases that are more common (e.g., Alzheimer's disease) or often manifest differently (e.g., myocardial infarction) in women are well-recognized, but why these differences occur, and what to do about them, is unclear. Moreover, women are more likely than men to report experiencing adverse events when taking medication, and the reasons for this, too, are mostly unknown.

One major reason given for this lack of knowledge is the history of underrepresentation of women in clinical research. This underrepresentation was not the result of women's explicit exclusion from participation in clinical trials. Rather, it was the result of federal policies enacted in response to the recognition of the damage wrought by thalidomide and diethylstilbestrol, both of which were prescribed to pregnant women without adequate evidence of their safety, and both of which resulted in significant harm to the children of these women. As a result of these harms, the US Food and Drug Administration (FDA) instituted a policy in 1977 that recommended excluding all women "of child-bearing potential" from early phase trials. This included Phase I trials, which are the first studies in humans, and which investigate how the medication being tested is metabolized and whether it is safe, and Phase II trials, which begin to examine effectiveness for the health condition to be treated, as well as the appropriate dosage and treatment frequency to achieve the desired clinical outcome(s). Notably, this policy also protected drug manufacturers from liability related to possible teratogenic effects of their products, which may go some way toward explaining why, in practice, this policy also led to the exclusion of women from later (Phase III) studies, in which they were still eligible to participate. This third phase of research provides the information required to support an application to the FDA for approval to market the drug; thus, drugs were being approved without sufficient evidence of their efficacy or safety in women.

From this perspective, the main patriarchal value influencing biomedical research is androcentrism, or the tendency to focus on men, and/or to take men to represent all humans (Little 1996). Critics have argued that researchers' focus on male participants has resulted in both a lack of knowledge about health needs specific to women and a failure to realize that health problems faced by both sexes may manifest differently in women. Moreover, they have pointed to an inconsistency in the reasoning for focusing only on males: One justification for using male animals was to avoid the complications and confounds caused by women's hormonal cycles. This suggests that females are different enough from males that they need to be considered separately. At the same time, however, researchers assumed that the results of studies conducted in males could readily be extrapolated to females, which could only be the case if differences between them were negligible. This inconsistency is the result of androcentrism, both in taking men to be the default and in seeing women's differences from men as distorting the evidence regarding a drug's "true" effect (Little 1996, 8).

Thanks in large part to women's health advocacy, things have changed a great deal. In 1986, the NIH introduced a policy that aimed to encourage the inclusion of women, and of racial and ethnic minority groups, in Phase III clinical research. Four years later, it released a report documenting the lack of progress toward greater inclusion (U.S. GAO 1990). Three years after that, Congress passed the NIH Revitalization Act, which mandated the inclusion of women and of racial and ethnic minority groups in all NIH-sponsored research, and in the same year, the FDA reversed its 1977 policy prohibiting women of childbearing potential from participating in early phase clinical studies. These federal efforts have all promoted what Steven Epstein (2007) calls the "inclusion-and-difference paradigm": Women must be included in biomedical research, and study results must be analyzed in such a way that outcome differences between women and men are identified. Since the early 1990s, there has been some progress in increasing women's enrollment in clinical trials, though it has been slower than many would have liked (Geller et al. 2018; Woitowich, Beery, and Woodruff 2020).

From the perspective of women’s health advocates, promoting the use of both female and male animals in preclinical research appears to be the next logical step in combating androcentrism in biomedical research. In an interview with Epstein, Sherry Sherman of the NIH Institute on Aging described her experience in graduate school during the 1980s. “[W]hen I began to do research in rats, I noticed that all rat studies were done on males” (Epstein 2007, 57). This has largely continued to be the case, and the use of male animals has been motivated by the same androcentric reasoning that shaped clinical research. A group of prominent researchers who support preclinical research on sex differences have described their colleagues’ worries that including female animals would double the sample size they needed and require them to control for the estrous cycle in female animals (McCarthy et al. 2012).

As with increasing women’s involvement in clinical research, change in preclinical research has been slow. The first publication to call for greater attention to possible sex differences in preclinical research was published in 2001; it was here that the mantra “every cell has a sex” was first introduced. (It also seems to be the source of phrasing that emphasizes “biological” sex differences, an idea that I will return to later in the chapter.) This report laid out the case for research on sex differences, reviewing what was then known about sex differences at the “cellular, developmental, organ, organismal, and behavioral levels” (Wizemann and Pardue 2001, ix). The authors concluded that “[t]here is now sufficient knowledge of the biological basis of sex differences to allow the generation of hypotheses with regard to health” (xix). Despite this conclusion, policy change was again slow to occur. The NIH policy on SABV was not introduced until 2014.² Like the 2001 report, the 2014 announcement of the new policies emphasized the link between preclinical and clinical research, saying that “[t]he overreliance on male animals and cells in preclinical research obscures key sex differences that could guide clinical studies” (Clayton and Collins 2014, 282). In addition, like those who argued for the inclusion of women in clinical research, Clayton and Collins argued that reliance on male animals in preclinical studies might result in harm to women.

Overall, the story for sex difference research, from this perspective, is one of improving women’s health through overcoming androcentrism in clinical research. Other feminist scholars, however, have a very different perspective. I will turn to this perspective in the next section.

40.2 Sex difference research and its feminist critics

Contrary to the view that research on sex differences is lacking, and should be encouraged, some feminist scholars have argued that there has actually been a great deal of research on sex differences and that it has been more likely to harm than to help women. Historically, this research has been used explicitly to support explicitly sexist views that aimed to buttress claims of inherent male superiority (Bluhm 2020), and even though this is no longer the case, gender stereotypes, rooted in essentialist views of sex and gender, continue to permeate sex difference research. These essentialist views reflect the idea, whether made explicit or left as a tacit assumption, that men and women are fundamentally different and that these differences are both innate and unchangeable.

Theories about sex differences long predate scientific inquiry (Dea 2016), but science has often been used to “prove” that women and men are different, and that these differences are natural and immutable. Historically, some of these theories were also intended to keep women in their proper place, by explaining their intellectual inferiority (Fee 1979; Tavris 1992) and even arguing that the attempt to compete academically with men would threaten a woman’s fertility or cause her to give birth to enfeebled children by shifting her energy from her reproductive organs to her brain (Sayers 1983). Other theories, while also viewing women as primarily reproductive creatures, claimed that both women’s physical characteristics and their temperaments suited them for domestic life. Their virtues were seen as complementary to men’s, and equally necessary for society to function (Laqueur 1992; Russett 1991).

This kind of theory exhibited what is now called benevolent sexism (Glick et al. 2000); women were celebrated even as they were firmly put in their place.

This kind of research exemplifies all three patriarchal values described by Longino. They are androcentric in viewing women as defined solely by their reproductive differences from men; they are sexist in “presupposing, asserting, or implying the inferiority of women, the legitimacy of their subordination, [and] the legitimacy of sex-based prescriptions of social roles and behaviors” (1990, 129) and they are essentialist in that they assume a “throughgoing dimorphism or sexual essentialism” that suggests that “‘they’ are made for and hence complementary to ‘us’” (Longino 1990).

I have previously argued that, in contemporary neuroscience research, the issue of sexism is more complicated (Bluhm 2020). Contemporary researchers do not say that women are inferior to men. They may explicitly state that the sexes are equal (Cahill 2014) or that their research doesn’t imply the superiority of one sex (Buss 1995). Yet many feminist critics have pointed out how deeply this research is informed by gender stereotypes, on which men are active, analytical, and aggressive, while women are passive, emotional, and nurturing. If anything, I believe, essentialism has become even more firmly entrenched, and tracing this process of entrenchment can help us to see how beliefs about women, men, and the differences between them also shape the call for more sex difference studies in preclinical research.

Establishing this claim requires a brief overview of the contemporary framework used by sex difference researchers, showing how the search for sex differences has developed and expanded over the past 70 years or so. During the middle decades of the twentieth century, much of the basic research on sex differences focused on the development of the reproductive organs and of the brain. The “Jost paradigm” was created to explain the development of sex organs in the fetus. Chromosomal sex (XY or XX) was held to be responsible for the development of the gonads (testes or ovaries), which produced sex hormones. More specifically, in males, the Y chromosome led to the development of the testes, which produced the “masculinizing” hormone, testosterone. Testosterone, in turn, caused the masculinization of the rest of the internal reproductive organs and the genitalia. In the absence of a Y chromosome, no testes were formed, and no testosterone produced, and the animal developed female organs. These findings gave rise to the longstanding view that female development was the “default” pathway, occurring in the absence of testosterone. The Jost paradigm thus connected the binary difference in sex chromosomes with the dimorphism of the reproductive system.

In 1959, this line of research was further extended to account for differences in sexual behavior. Charles Phoenix and his colleagues injected pregnant guinea pigs with testosterone and found that, not only were their female offspring born with “masculinized” genitalia, but when the animals were adults, another injection of testosterone caused them to engage in male-typical sexual behaviors (e.g., mounting) more frequently than did female animals not exposed to prenatal testosterone. In addition, the animals were injected with estrogen and progesterone (the “female” sex hormones). While both testosterone-exposed and -unexposed animals then displayed female-typical sexual behaviors (lordosis), this was less frequent in the testosterone-exposed animals. These findings led to the development of the “organizational-activational hypothesis,” which proposes that prenatal testosterone masculinizes the brain, so that the increase of testosterone levels at puberty acts on a brain that was already primed for male sexual behavior. By contrast, as in the Jost paradigm, female animals follow the “default” developmental pathway; in the absence of testosterone, there is no masculinization and the brain remains feminized (Phoenix et al. 1959).

Yet note the phrasing in the previous paragraph related to the frequency of sex-typical behaviors. Reproductive organs are sexually dimorphic in rodents (and humans)—in the vast majority of cases, reproductive organs develop into one of two, male or female, forms. Things are not so simple with sexual behaviors, however, as even “normal” male and female animals will engage in the sexual behaviors associated with the other sex. As it happens, this kind of overlap is more common with many

characteristics than a dimorphism of two distinct forms. One way to think about this is to say that most characteristics that differ between males and females are *not* dimorphic; those characteristics are better understood as occurring along a continuum, on which average values for males and females differ, and there is some (or a lot of) overlap between the sexes. But another way to think about this is to change the definition of dimorphism, which is what has actually tended to happen among sex difference researchers. In a chapter on sexually dimorphic behaviors, Goy and McEwan note these two alternatives and explicitly opt for the second: In their review “the term sexually dimorphic does not usually indicate the presence of a given response in one sex exclusively and its absence in the other, but, rather, *any measurable difference in the parameters of the response for the two sexes*” (1980/2007, 7, italics added).

It is likely that this redefinition of dimorphism resulted from an underlying belief in sex essentialism, and it has also reinforced it. Similarly, as Rebecca Jordan-Young has noted, the organizational-activational hypothesis has moved over time from being a (potentially falsifiable) theory to being viewed as a “‘background fact’ of development” (2010, 37). As the next two sections will show, the scope of the organizational-activational framework has also expanded in two ways, each of which also derives its plausibility from and reinforces sex essentialism.

40.3 Behavior beyond reproduction

More than 60 years after it was first proposed, the organizational-activational hypothesis still plays a central role in research on sex differences, but it has expanded well beyond differences related directly to reproduction. This expansion has had two directions. First, starting in the 1970s, researchers working with children who had been exposed to abnormal amounts of sex hormones during fetal development (either because of a disorder or because their mothers were prescribed hormone treatments while pregnant) became interested in the behavioral effects of such hormonal exposure. Much of this research was conducted in girls and women with congenital adrenal hyperplasia (CAH), a condition that results in exposure to high levels of (“masculinizing”) androgens during fetal development. John Money and Anke Ehrhardt (1972) investigated whether these individuals were more likely than their sisters, who did not have CAH, to have masculine behaviors and interests. Depending on the age of their participants, these masculine tendencies might be manifested in engaging in rough-and-tumble play, failing to have fantasies about their future wedding, or interest in a career and a lack of interest in motherhood. Later studies also expanded the range of masculine and feminine characteristics to include cognitive abilities such as spatial reasoning and mathematical ability, as well as verbal ability. (Research also focused on the amount of same-sex attraction in women with CAH, reflecting Longino’s claim that sex essentialism is also heterosexism.)

Psychology researchers of the time also investigated sex differences in samples of individuals with no known history of abnormal exposure to sex hormones in utero. Their studies also examined cognitive abilities, as well as perception, learning, personality, and social behavior (reviewed in [Maccoby and Jacklin 1974](#)). Neuroscientists of that era looked for sex differences in the size and shape of various parts of the brain (notably the corpus callosum). A decade later, the increasing availability of neuroimaging technologies in the following decade, led to studies that aimed to find sex differences in brain function associated with tasks and abilities thought to be different between women and men. All of these areas of research have been criticized by feminist scholars who pointed out the way in which gender stereotypes influenced the design, analysis, and interpretation of these studies.³

One way to push back against these stereotypes is to emphasize the extent to which behavioral and psychological differences are the result of socialization and gendered expectations. While there

is good reason to think that this is indeed the case, many researchers seem to think that social influences have only a limited ability to influence innate differences. For example, the neuroscientist Larry Cahill (2014) has argued that, while experience can modify the brain (a phenomenon known as neural plasticity), “[t]he brain is plastic only within the limits set by biology.” It is clear, too, that “biological” here means “innate”—Cahill continues,

[t]he idea that we should use the brain’s plasticity to work against inborn masculine or feminine predispositions in the brains of children is as ill conceived as the idea that we should encourage left-handed children to use their right hand.

Similarly, Margaret McCarthy and Gregory Ball describe a feminist critic of sex difference research as coming to a greater acceptance over time of “the biological basis of human brain differences” (2011, 1) and they refer in the next paragraph to “the neural basis of sex differences in the brain” (McCarthy and Ball 2011). But the brain *is* biological, and it does not become less so if it changes in response to social experiences. As I have previously noted, this conflation of “biological” and “innate” predisposes researchers to think about sex differences in the brain as largely “natural,” and not a result of gendered differences in experience (Bluhm 2020).

40.4 Sex everywhere?

Not only has the search for sex differences been extended to psychological and behavioral characteristics unrelated to reproduction, but it has also expanded to parts of the body not directly related to reproduction. Even before the NIH SABV policy brought greater attention to sex differences in basic research, searching Google Scholar reveals that numerous researchers in cardiology, endocrinology, pain research, immunology, digestive physiology, and a host of other areas had begun citing the 2001 report by Wizemann and Pardue in their publications.

This research is also informed by the organizational-activational framework. In an article published to mark the 50th anniversary of the Phoenix et al. paper that introduced the framework, the biologist Arthur Arnold describes this research as “the foundation for a unified theory of sexual differentiation of all mammalian tissues” (Arnold 2009, 570). But he also says that hormones are not the only cause of sex differences; researchers now believe that a full understanding of sexual differentiation also requires taking into account the role of genes located on the (X and Y) sex chromosomes. Arnold and a colleague have therefore urged researchers to study the “sexome,” which they defined as “the sum of all sex-specific and sex-biased modulatory interactions that operate within [gene] networks” (Arnold and Lusi 2012, 2551).

This second expansion of the organizational-activational framework means that sex differences are everywhere. Not only does every cell have a sex, and all mammalian tissues, but so, apparently, does every organ. Published articles refer to the sex of livers, lungs, hearts, kidneys, brains and shoulders, knees and foot bones (Conley, Rosenberg, and Crowninshield 2007; Ferrari et al. 2004; Harris et al. 2018; Knowlton and Korzick 2014; Luders and Kurth 2020; McFarlane and Heneghan 2004; Pignataro et al. 2017; Wang et al. 2021). To these scientists, referring to “the female lung” is apparently as natural as (and, possibly, more scientific than) referring to “women’s lungs.” Moreover, given the tendency I described above to equate “biological” with “natural” and “innate,” this way of talking reinforces sex essentialism: Women and men are simply different, anywhere it is possible to have a difference.

But what does it even mean to say that a lung is female or male, or that a piece of liver tissue has a sex? I will turn to that question in the next, and final, section.

40.5 What even is sex?

The feminist neuroscientist Daphna Joel suggests that the term “sex,” as used in the biomedical literature, is problematically ambiguous. It refers both to a demographic distinction between male and female organisms—a distinction made (usually) on the basis of the appearance of the genitals—and to what Joel describes as a biological “system that affects other systems (e.g., genitalia, brains)” (2016, 336). She further says that “[a]s a system, it is currently widely accepted that sex includes a genetic and a hormonal component” (Joel 2016). This latter sense, sex as a system, is consistent with the research I have just described and, more broadly, with the whole idea of SABV.

Joel notes that the demographic categories of male and female are based on what she calls “3G” sex, referring to their basis on genes, gonads, and genitalia (Joel 2012). These characteristics are (mostly) dimorphic and (nearly) dichotomous, so that most people who have the usual “female” form of one characteristic almost certainly have the “female” form of the other two. The same cannot be said for “sex as a system” of genetic and hormonal processes, for which there is a significant amount of overlap in any characteristic studied. Joel suggests that it is the ambiguity of the term “sex” that makes it seem natural to assume that “sex [is] a dichotomous system or process that exerts profound effects on other systems (e.g., the brain), leading to sexual dimorphism also of these systems” (2016, 337).

I think that Joel is correct about this. But in fact, even the idea that sex is a biological system is a product of how deeply sex essentialism has taken root in contemporary research. Yet, it is possible to accept that human beings (generally) fall clearly into the category of female or of male, without also thinking that they got that way because of “sex”—the process is not the product. The biologist Frank Lillie, in 1939, made exactly this claim:

There is no such biological entity as sex. What exists in nature is a dimorphism within species into male and female individuals...Sex is not a force that produces these contrasts; it is merely a name for our total impression of the differences. (3)

He acknowledged that (even then) the terms “male” and “female” were expanding from descriptions of an individual based on the type of gametes (sperm or eggs) they produced and had “come to be applied also to the gametes themselves, to the determiners of male and female characteristics, and to the zygote destined to produce a male or a female” (Lillie 1939). I have shown that this mission creep of sex has become ever more entrenched as research on sex differences has expanded. We seem to have moved from sex as “a total impression of the differences” between females and males, to an impression of total difference between them.

Moreover, this expansion may well be exacerbated by feminists’ attempts to improve our understanding of women’s health. Here, too, there is historical precedent. Marianne van den Wijngaard has argued that, during the 1970s, second-wave feminism “reinforced the biological dichotomization of female and male behavior in research in animals and humans” (1994, 63). The earliest animal research on sexual behavior focused on the behavior of male animals (and “masculinized” female animals), with male mating behavior distinguished into a number of distinct aspects, while female behavior was understood simply as responsiveness (signaled by lordosis) to male mounting behavior. Thus, “[a]ctive female sexual behavior did not exist within the framework of the organization theory in the 1960s...androgens and masculinity ruled the investigations of behavior” (van den Wijngaard 1994, 64). As feminists began to demand equality, and more women began to enter the field of behavioral neuroendocrinology, more attention was paid to female reproduction, with papers that called for more research with female animals explicitly adopting the rhetoric of the Women’s Liberation Movement (van den Wijngaard, 68).

However, van den Wijngaard shows that the uptake of research on female animals depended on the extent to which it was compatible with brain organization theory. Both the behavior of male and female animals, and the differences between them, had to be explained in terms of prenatal hormones. Female behavior, it seems, could be studied as long as it fits into the prevailing ideas about the nature and cause of sex differences. In short, feminist aims to redress androcentrism ended up buttressing sex essentialism.

I have shown in this chapter that the contemporary push to investigate sex differences in pre-clinical research similarly reflects and reinforces sex essentialism. The slogans describing sex as a “biological variable” and every cell as having a sex have, I think, received the uptake that they have precisely because they fit with existing basic research on sex differences. My analysis suggests that sex essentialism may prove to be as big an impediment to women’s health as the androcentrism it has replaced.

I want to conclude, however, by briefly considering whether it is possible to avoid this outcome and to use biological research incorporating a search for sex differences to improve women’s health without falling prey to sex essentialism. Feminist critics of the SABV approach have argued that it is likely to draw our attention away from gendered social influences on health, and to prevent the development of complex models of interactions between biological processes and social factors (Richardson et al. 2015; Springer, Mager Stellman, and Jordan-Young 2012), but they have also emphasized that a commitment to biology does not necessitate accepting sex essentialism. For example, Stacey Ritz has emphasized that “sex” is not a (single) biological variable, but “a complex collection of characteristics across multiple levels of organization” (Ritz 2017, 319), and has discussed work from her own lab that examined the effects of estrogen on T-cell differentiation *in vitro*, while attempting “to avoid the inference that estrogen was in some way a proxy for femaleness” (Ritz 2017, 323). Because estrogen “is found in all human bodies and is produced through the integration of genetic, cellular, physiological, social, and environmental signals” (Ritz 2017, 323), she suggests, estrogen levels, and their effects on other biological processes, can help us to recognize the complexity of biology and its interaction with social factors. From this perspective, developing a better understanding of biology is likely to be our best way to avoid sex essentialism.

Notes

- 1 The website also acknowledges that health disparities occur among women due to factors such as race/ethnicity, geography, and age. Given my focus in this chapter on the notion of sex as a biological variable, I will not address these important points. <https://swhr.org/programs/womens-health-equity-initiative/>.
- 2 The policy does not explain what is meant by “sex as a biological variable”; my best guess is that it is intended to both emphasize the pervasive nature of sex differences in the body and to distinguish (biological) sex differences from (social) gender differences. See, for example, the introductory paragraph in Clayton (2018).
- 3 For example, gender stereotypes (e.g., women are more emotional; men are less talkative) have influenced both the choice of characteristics to examine for potential sex differences and the interpretation of results (whether or not sex differences are actually found). In addition, researchers may change how they measure these characteristics and analyze their data in order to find sex differences. (See, for example, Fausto-Sterling (1992, 2000); Fine (2010, 2017); Jordan-Young (2010); and a number of the papers in Bluhm et al. (2012)).

References

- Arnold, Arthur P. 2009. “The Organizational-Activational Hypothesis as the Foundation for a Unified Theory of Sexual Differentiation of All Mammalian Tissues.” *Hormones and Behavior* 55: 570–78. <https://doi.org/10.1016/j.yhbeh.2009.03.011>.
- Arnold, Arthur P., and Aldons J. Lusis. 2012. “Understanding the Sexome: Measuring and Reporting Sex Differences in Gene Systems.” *Endocrinology* 153 (6): 2551–55. <https://doi.org/10.1210/en.2011-2134>.

- Bluhm, Robyn. 2020. “Neurosexism and Our Understanding of Sex Differences in the Brain.” In *The Routledge Handbook of Feminist Philosophy of Science*, edited by Sharon Crasnow and Kristen Intemann. Routledge.
- Bluhm, Robyn, Heidi L. Maibom, and Anne Jaap Jacobson, eds. 2012. *Neurofeminism: Issues at the Intersection of Feminist Theory and Cognitive Science*. Palgrave.
- Buss, David M. 1995. “Psychological Sex Differences: Origins through Natural Selection.” *American Psychologist* 50 (3): 164–68. <https://psycnet.apa.org/doi/10.1037/0003-066X.50.3.164>.
- Cahill, Larry. 2014. “Equal ≠ Same: Sex Differences in the Human Brain.” *Cerebrum*. http://www.dana.org/Cerebrum/2014/Equal_%E2%89%A0_The_Same__Sex_Differences_in_the_Human_Brain/.
- Clayton, Janine A. 2018. “Applying the New SABV (Sex as a Biological Variable) Policy to Research and Clinical Care.” *Physiology & Behavior* 187: 2–5. <https://doi.org/10.1016/j.physbeh.2017.08.012>.
- Clayton, Janine A., and Francis S. Collins. 2014. “Policy: NIH to Balance Sex in Cell and Animal Studies.” *Nature* 509 (7500): 282–83. <https://doi.org/10.1038/509282a>.
- Conley, Sheryl, Aaron Rosenberg, and Roy Crowninshield. 2007. “The Female Knee: Anatomic Variations.” *Journal of the American Academy of Orthopaedic Surgeons* 15: S31–36.
- Dea, Shannon. 2016. *Beyond the Binary: Thinking about Sex and Gender*. Broadview Press.
- Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. University of Chicago Press.
- Fausto-Sterling, Anne. 1992. *Myths of Gender: Biological Theories about Women and Men*. Revised ed. Basic Books.
- Fausto-Sterling, Anne. 2000. *Sexing the Body: Gender Politics and the Construction of Sexuality*. Basic Books.
- Fee, Elizabeth. 1979. “Nineteenth-Century Craniology: The Study of the Female Skull.” *Bulletin of the History of Medicine* 53 (3): 415–33. <https://www.jstor.org/stable/44450930>.
- Ferrari, Jill, David A. Hopkinson, and Alf Linney. 2004. “Size and Shape Differences Between Male and Female Foot Bones.” *Journal of the American Podiatric Medical Association* 94 (5): 434–52. <https://doi.org/10.7547/0940434>.
- Fine, Cordelia. 2010. *Delusions of Gender: How our Minds, Society, and Neurosexism Create Difference*. W.W. Norton & Company.
- Fine, Cordelia. 2017. *Testosterone Rex: Myths of Sex, Science, and Society*. W.W. Norton & Company.
- Geller, Stacie E., Abigail R. Koch, Pamela Roesch, Amarett Filut, Emily Hallgren, and Molly Carnes. 2018. “The More Things Change, the More They Stay the Same: A Study to Evaluate Compliance With Inclusion and Assessment of Women and Minorities in Randomized Controlled Trials.” *Academic Medicine* 93 (4): 630–35. <https://doi.org/10.1097/ACM.0000000000002027>.
- Glick, Peter, Susan T. Fiske, Antonio Mladinic, et al. 2000. “Beyond Prejudice as Simple Antipathy: Hostile and Benevolent Sexism across Cultures.” *Journal of Personality and Social Psychology* 79 (5): 763–75. <https://doi.org/10.1037/0022-3514.79.5.763>.
- Goy, Robert W., and Bruce S. McEwan. 1980/2007. “Sexually Dimorphic Behavior: Definition and the Organizational Hypothesis.” *Reprinted in Sex and the Brain*, edited by Gillian Einstein. MIT Press.
- Harris, Autumn N., Hyun-Wook Lee, Gunars Osus, et al. 2018. “Differences in Renal Ammonia Metabolism in Male and Female Kidney.” *American Journal of Physiology—Renal Physiology* 315 (2): F211–22. <https://doi.org/10.1152/ajprenal.00084.2018>.
- Joel, Daphna. 2012. “Genetic-Gonadal-Genitals Sex (3G-Sex) and the Misconception of Brain and Gender, or, Why 3G-Males and 3G-Females Have Intersex Brain and Intersex Gender.” *Biology of Sex Differences* 3: 27. <https://doi.org/10.1186/2042-6410-3-27>.
- Joel, Daphna. 2016. “Captured in Terminology: Sex, Sex Categories, and Sex Differences.” *Feminism & Psychology* 26 (3): 335–45. <https://doi.org/10.1177/0959353516645367>.
- Jordan-Young, Rebecca M. 2010. *Brain Storm: The Flaws in the Science of Sex Differences*. Harvard University Press.
- Knowlton, Anne A., and Donna H. Korzick. 2014. “Estrogen and the Female Heart.” *Molecular and Cellular Endocrinology* 389 (1–2): 31–39. <https://doi.org/10.1016/j.mce.2014.01.002>.
- Laqueur, Thomas. 1992. *Making Sex: Body and Gender from the Greeks to Freud*. Harvard University Press.
- Lillie, Frank R. 1939. “General Biological Introduction.” In *Sex and Internal Secretions: A Survey of Recent Research*, edited by Edgar Allen Williams and Wilkins Co. 2nd ed. <https://psycnet.apa.org/doi/10.1037/13343-001>.
- Little, Margaret Olivia. 1996. “Why a Feminist Approach to Bioethics?” *Kennedy Institute of Ethics Journal* 6 (1): 1–18. <https://doi.org/10.1353/ken.1996.0005>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press.

- Luders, Eileen, and Florian Kurth. 2020. "Structural Differences between Male and Female Brains." *Handbook of Clinical Neurology* 175: 3–11. <https://doi.org/10.1016/B978-0-444-64123-6.00001-1>.
- Maccoby, Eleanor E., and Carol N. Jacklin. 1974. *The Psychology of Sex Differences*. Vol. 1. Stanford University Press.
- McCarthy, Margaret M., Arthur P. Arnold, Gregory F. Ball, Jeffrey D. Blaustein, and Geert J. DeVries. "Sex Differences in the Brain: The Not So Inconvenient Truth." *Journal of Neuroscience* 32 (7): 2241–47. <https://doi.org/10.1523/JNEUROSCI.5372-11.2012>.
- McCarthy, Margaret M., and Gregory F. Ball. 2011. "Tempests and Tales: Challenges to the Study of Sex Differences in the Brain." *Biology of Sex Differences* 2 (4): 1–4. <https://doi.org/10.1186/2042-6410-2-4>.
- McFarlane, Ian G., and Michael A. Heneghan. 2004. "Autoimmunity and the Female Liver." *Hepatology Research* 28 (4): 171–76. <https://doi.org/10.1016/j.hepres.2003.12.002>.
- Money, John, and Anke A. Ehrhardt. 1972. *Man and Woman, Boy and Girl: Gender Identity from Conception to Maturity*. Johns Hopkins University Press.
- Phoenix, Charles H., Robert W. Goy, Arnold A. Gerall, and William C. Young. 1959. "Organizing Action of Prenatally Administered Testosterone Propionate on the Tissues Mediating Mating Behavior in the Female Guinea Pig." *Endocrinology* 65: 369–82. <https://doi.org/10.1210/endo-65-3-369>.
- Pignataro, Francesca S., Matteo Bonini, Alessandra Forgone, Serena Melandri, and Omar S. Usmani. 2017. "Asthma and Gender: the Female Lung." *Pharmacological Research* 119: 384–90. <https://doi.org/10.1016/j.phrs.2017.02.017>.
- Richardson, Sarah S., Meredith Reiches, Heather Shattuck-Heidorn, Michelle Lynne LaBonte, and Theresa Consoli. 2015. "Focus on Preclinical Sex Differences Will Not Address Women's and Men's Health Disparities." *Proceedings of the National Academy of Sciences of the United States of America* 112 (44): 13419–20. <https://doi.org/10.1073/pnas.1516958112>.
- Ritz, Stacey. 2017. "Complexities of Addressing Sex in Cell Culture Research." *Signs: Journal of Women in Culture and Society* 42 (2): 307–27.
- Russett, Cynthia E. 1991. *Sexual Science: The Victorian Construction of Womanhood*. Harvard University Press.
- Sayers, Janet. 1983. *Biological Politics: Feminist and Anti-Feminist Perspectives*. Tavistock Publications.
- Springer, Kristen W., Jeanne Mager Stellman, and Rebecca M. Jordan-Young. 2012. "Beyond a Catalogue of Differences: A Theoretical Frame and Good Practice Guidelines for Researching Sex/Gender in Human Health." *Social Science & Medicine* 74 (11): 1817–24. <https://doi.org/10.1016/j.socscimed.2011.05.033>.
- Tavris, Carol. 1992. *The Mismeasure of Woman*. Simon and Shuster.
- U.S. General Accounting Office. 1990. "Statement of Mark Nadel, National Institutes of Health: Problems in Implementing Policy on Women in Study Populations." National Institutes of Health. June 18. <https://www.gao.gov/assets/t-hrd-90-50.pdf>.
- van den Wijngaard, Marianne. 1994. "Feminism and the Biological Construction of Female and Male Behavior." *Journal of the History of Biology* 27 (1): 61–90. <https://www.jstor.org/stable/4331296>.
- Wang, Jun, I-Lin Wang, Rui Hu, et al. 2021. "Intermediate Effects of Acupuncture on Explosive Force Production and Stiffness in Male Knee Joint." *International Journal of Environmental Research and Public Health* 18 (18): 9518. <https://doi.org/10.3390/ijerph18189518>.
- Wizemann, Theresa M., and Mary-Lou Pardue. 2001. *Exploring the Biological Contributions to Human Health: Does Sex Matter?* National Academy Press.
- Woitowich, Nicole C., Annaliese Beery, and T. Teresa Woodruff. 2020. "Meta-Research: A 10-Year Follow-Up Study of Sex Inclusion in the Biological Sciences." *eLife* 9: e56344. <https://doi.org/10.7554/eLife.56344>.

VALUES IN HUMAN GENOMICS

Yasmin Haddad and Celso Neto

41.1 Introduction

Genomics is a discipline that focuses on the study of an organism's genetic material, relying on sequencing technologies that allow a deep understanding of genetic interactions and their effects. Genomics is comprehensive: it seeks to study the whole genome of living things, allowing scientists to gain insights into complex biological systems across various taxonomic groups. In this chapter, we will focus on *human* genomics—a particularly promising field of study that enhances our understanding of the origin of complex diseases, their underlying mechanisms, patterns of inheritance, and the discovery of new genes. Human genomics can also improve our understanding of human evolutionary history, influencing our notions of identity and belonging. Thus, this domain of research has significant implications for fields such as medicine, biotechnology, anthropology, pharmacogenomics, and evolutionary biology.

Human genomics is a fast-evolving area due to advancements in data collection technologies and statistical methods of data clustering. For example, genome-wide association studies (GWAS) proceed by selecting study populations, large-scale genotyping, data processing, and testing for associations between genetic markers and traits (Uffelmann et al. 2021). These associations supposedly tell scientists about the genetic basis of phenotypic traits (e.g., a rare disease or the origin of complex diseases). In this research context, it is common practice to divide large populations into sub-populations to establish comparisons across groups. One standard practice is to label sub-populations using population descriptors such as “race,” “ethnicity,” or “ancestry.” These labels are created both from self-identifying information at the stage of participant recruitment but also in alignment with clustering techniques that aim to identify genetic similarities and differences between and within those populations. While human genetic variation is mostly continuous, scientists apply clustering techniques such as PCA, STRUCTURE, and UMAP to simplify large data sets and to group human samples according to (typically small) patterns of genetic similarity (National Institute of Health 2020; Novembre 2016). Genetically similar individuals within a sample are thus labelled with the goal of facilitating comparisons across groups. However, labels are frequently inconsistently used without recognition of potential racist implications that may result from slippage between concepts such as “ancestry” and “race” (Popejoy 2021).

While advancing our understanding of genetics and health outcomes, the field of human genomics has also grappled with significant racist legacies. Early genetic studies often operated based on

misconceptions about human differences that were used to justify racial hierarchies (Borrell et al. 2021; Kampourakis and Peterson 2023; Reardon 2005). The historical misuse of genomics has contributed to the reification of racial categories and the reinforcement of stereotypes that persists to this day (Bolnick 2008; Gannett 2004; Popejoy 2021). Moreover, lack of diversity has been an ongoing problem in genomic studies, which disproportionately rely on data from individuals of European descent (Hindorff et al. 2018). This bias not only undermines the scientific validity of genomics but also exacerbates healthcare disparities by failing to account for genetic diversity across all human populations.

Today, several efforts are put in place to ensure that racist legacies are no longer a part of human genomics and to ensure that genetic diversity is more accurately represented in large-scale studies. For instance, international projects such as H3Africa and 3MAG seek a better representation of African genomes, as they understand that a more equitable representation of genetic diversity can help achieving more equitable health outcomes (Wonkam 2021). Moreover, scientific institutions such as the National Academies of Science, Engineering and Medicine (NASEM) work to design policies for the responsible use of “race,” “ethnicity,” and “ancestry” in human genomics research that avoids perpetuating racism (National Academies of Science, Engineering and Medicine 2023). Such efforts are evidence that human genomics is not *value-free*. In other words, human genomics is embedded in social contexts and can be influenced by the views and preferences that exist in those contexts. Some of these values (e.g., racism) might be undesirable, while other values (e.g., the promotion of equitable health outcomes) might be desirable. Hence, geneticists and scientific institutions constantly adapt their practices to either avoid or promote certain values.

In this chapter, we will discuss the influence of values in human genomics. Our goal is to show how the influence of values has important consequences both in theory and practice, deserving the attention from philosophers but also scientists and policymakers. First, we will examine and reject the idea that human genomics should be value-free (Section 41.2). Second, we will consider in what ways values should influence genomics (Section 41.3). Finally, we discuss how these concerns arise in one of the largest current scientific endeavors in the field of human genomics: the *All of Us* cohort study sponsored by the National Institute of Health (henceforth NIH) (Section 41.4). Ultimately, this chapter illustrates how lack of attention to values can lead us astray.

41.2 Should human genomics be value-free?

Like any other science, human genomics is subject to strict standards and methodologies with the goal of being empirically adequate. However, it is also prone to the influence of preferences of scientists, scientific institutions, and other societal actors. In the philosophy of science, we call these preferences “values” because they state whether a certain outcome is desirable or not (Elliott and Korf 2024). For instance, the value of anti-racism reflects the belief that racism is unacceptable, which can drive scientific institutions to avoid using racial terminology in genomics research (National Academies of Science, Engineering, and Medicine 2023).

Values can influence the overall ecosystem of scientific research. This influence might happen across different stages of the scientific process, from approaching funders, to designing research questions and methodologies, up to deciding how to disseminate the results (Elliott 2022). Preferences might inform the decisions made by scientists at each of these stages. One particularly important stage in most genomics studies is the choice of how to classify human populations. While philosophical literature suggests that scientific classification is a value-laden task, the value-ladenness of genetic classifications is understudied (Ahn 2020; Ludwig 2016; Reydon and Ereshefsky 2022). In human genomics, some classification systems are thought to be more appropriate than others if they map more accurately onto actual population structures (Rosenberg et al. 2002). As such, geneticists might

be tempted to think that accurate classification systems reflect a biological reality and are therefore not prone to the influence of values. One might think that preferences do not (and should not) influence successful scientific classifications. However, to what extent is this a justified assumption? In this section, we will consider whether human genomics should aspire to be value-free, with a special focus on the task of human classification into different groups.

To address this issue, it might be useful to start with the distinction between epistemic and non-epistemic values. Epistemic values are those that contribute to the scientific goal of producing reliable information and epistemically acceptable scientific practices. Examples of epistemic values include empirical accuracy, validity, replicability, and precision. These values are considered desirable guiding principles or systematic indicators of knowledge and signs of “empirically good” science (Elliott 2022). In contrast, non-epistemic values are personal, social, and cultural values that shape individual or group preferences about what science ought to be and how it ought to proceed (Longino 1990).¹ These values might be desirable or not; however, they are not meant to *systematically* indicate or contribute to knowledge in science. For instance, the non-epistemic values of social justice and anti-racism might rightfully guide many research projects, but such projects will not necessarily generate better (more reliable) information simply because of that.

It is a tacit consensus in the literature that epistemic values can play a desirable and key role in shaping scientific practices (Douglas 2009). When deciding between classificatory systems, for instance, scientists will rightly choose the ones that are more empirically adequate and methodologically sound. The more difficult question to be addressed is whether science and scientific classification should be free from *non*-epistemic values. This is the question that concerns many scientists and philosophers of science.

At first glance, it may seem that science should be free from non-epistemic values. If the goal of science is to produce knowledge claims, non-epistemic values might be irrelevant and sometimes even undesirable for this goal. Values can bias scientific research, making it a mere instrument of social and political agendas (Elliott 2017). This worry is often described as the problem of “wishful thinking,” namely the problem of accepting or rejecting a theory just because one wants it to be true or false (Anderson 2004; Brown 2013). This problem is not new to research on human classifications. For instance, what scholars call “scientific racism” can be understood as the attempt to vindicate racial classifications as if they were empirically true when in fact they were largely driven by racist values (Gould 1993). As this example illustrates, non-epistemic values might lead scientists to falsehood. The value-free ideal holds that science should strive to be free from such influences, thus avoiding problems such as “wishful thinking” and ensuring that science remains true to facts and evidence.

However, pursuing value freedom in human genomics can be problematic for two main reasons we will explore below: (i) slippage and (ii) underdetermination.

41.2.1 Slippage

Historically, the use of “race” in scientific contexts dangerously encouraged the idea that distinct human races do exist in a biological sense (Yudell 2011). Currently, scientists and scientific institutions are aware of this danger. Biological human races do not exist, and the term “race” is increasingly less present in scientific literature, particularly in human genomics (Byeon et al. 2021). However, other labels (population descriptors) are still applied to human populations, as in the case of “ethnicity” and “ancestry.” The latter has gained immense popularity in recent years and is often assumed to be a feature of the biological world (Dauda et al. 2023; Gannett 2014; Harden 2021). Trying to pursue value freedom, in this case, is related to a naïve conception of objectivity, which assumes that population descriptors used in genomics refer to biological groups that exist independently of human scientific practices. For example, Rosenberg et al. (2002) have famously outlined the genetic structure of

human populations, reaching the conclusion that part of this genetic structure maps onto continental ancestry. In other words, while most human variation is continuous, humans can still be meaningfully divided into ancestry groups based on patterns of genetic similarity. While some take this as evidence that labels in genomics do refer to biological groups that exist independently of human scientific practices, this inference is not straightforward. In fact, there was a complex decision-making process that involved conscious decisions on behalf of scientists to represent a number of clusters based on the number of existing continents (Bolnick 2008). Thus, value freedom encourages a simplistic view of human classification that dismisses the role of human decisions and preferences in the development of those classifications (Haddad 2025).

This simplistic view becomes especially problematic when considering the linguistic slippage between “race” and “ancestry.” While scientists might insist on the difference between race and ancestry, these terminologies overlap and are very easily conflated. For instance, continental ancestry labeling (e.g., “African,” “Asian,” and “European”) mirrors traditional classifications of human races popularized since the eighteenth century and still present in the US Census (Spencer 2014). The worry is that the language of ancestry reinforces racial beliefs and classification (Lewis et al. 2022). The value-free acceptance of human biological divisions based on ancestry might prompt implicit beliefs in the biological reality of races. Moreover, the slippage between “race” and “ancestry” has also been identified as deliberate ambiguity by sociologists of science, possibly indicating that the choice and use of population descriptors such as “race,” “ancestry,” and “ethnicity” is motivated by non-epistemic as well as epistemic values. Panofsky and Bliss (2017) show that the most important reason for choosing continental ancestry labeling is to establish connections between genetics and their extra-scientific contexts, such as researchers in the field, journals, and grant agencies. The idea is that ambiguous labels are more likely to be better understood both by study participants (who will more easily volunteer) and by the overall scientific community who sees benefits in using coarse-grained labels as they are more generalizable across studies. Thus, ethnographic analysis shows that there are different social considerations guiding the use of population descriptors and the construction of human classifications, indicating the value-ladenness of genomics. Ultimately, failure to acknowledge this value-ladenness can reinforce a simplistic view of human classifications and unwarranted racial beliefs.

41.2.2 Underdetermination

Another concern raised by philosophers of science is the problem of underdetermination (Anderson 2004). Broadly, underdetermination means that data itself is insufficient to either confirm or reject a given theory or hypothesis. For instance, when Rosenberg et al. (2002) use genomic data to infer that human genetic diversity maps onto continental ancestry classifications, they cannot reach this conclusion by simply relying on data. Background assumptions are also required to make the inference. There are numerous decisions involved, including why researchers choose to use “ancestry” as a population descriptor rather than “race” or “ethnicity.” Many of these decisions involve non-epistemic values motivated by extra-scientific factors, as exemplified in the previous paragraphs. In sum, if underdetermination means that data are insufficient to confirm or nullify a given hypothesis, then scientists must appeal to background assumptions to establish what data counts as evidence for the hypothesis and these assumptions are inevitably value-laden (Longino 2002, 2019).

The problems of slippage (i) and underdetermination (ii) indicate that human genomics is inevitably value-laden. With respect to underdetermination, one may argue here that scientists should therefore refrain from drawing conclusions until the evidence is unambiguous. However, it is just impractical and unrealistic to expect scientists to refrain from drawing conclusions, despite ambiguous evidence (Biddle 2013). Upon a closer look, (i) and (ii) also suggest that human genomics *should*

be infused by values. Slippage entails ethical risks, i.e., it creates the possibility of unintended harm. One example would be the use of “ancestry” in ways that reinforce the belief that biological races exist, a belief that has been traditionally weaponized against marginalized groups. If slippage is a constant possibility and entails ethical risks, scientists need to evaluate these risks and adopt values that can direct research toward minimizing those risks (Biddle and Kukla 2017; Douglas 2003). In fact, many scientists already do that, tacitly recognizing the importance of values in human genomics (Lewis et al. 2022; National Academies of Science, Engineering and Medicine 2023). Thus, aiming for the value-freedom of human genomics can be both unrealistic and dangerous. This attitude would likely reduce transparency and enable dubious values to influence that science.

Adopting values can be highly desirable. Values can set important research aims and be paramount to their pursuit (Elliott and McKaughan 2014). For instance, genomic projects like H3Africa and All of Us aim to understand and address health inequalities across populations. These projects can have a high social impact precisely because they recognize the value of increasing the representation of genomic diversity across underrepresented populations. Moreover, the projects are driven by the assumption about equity and beneficence, namely the idea that diverse human populations should equally benefit from human genomics research. Thus, values can shape research toward desirable outcomes and ensure that genomics research is conducted responsibly. However, if values are both inevitable and desirable, scientists have to face a key question: What kind of influence of non-epistemic values is acceptable and even desirable in human genomics?

41.3 What roles should values play in human genomics?

Philosophers of science have provided numerous rich characterizations of the different roles that values can play in science. In scientific practice, values are explicitly mentioned as worthy goals to be pursued to achieve empirically adequate science. One example is the recently published report issued by the National Academies of Sciences on the use of population descriptors in genomics (2023). This report reflects the consensus of an extensive collaboration among geneticists, social scientists, bioethicists, and other stakeholders in the field. The report specifically mentions “guiding principles” intended to “foster the highest ethical and empirical practices and to support trustworthy research to inform its recommendations” (2023, 95) that must guide and inform scientific practice in genomics. These guiding principles should influence genomics research throughout different research stages. Examples of such guiding principles—or values—include respect, beneficence, equity, justice (non-epistemic values) and validity, reproducibility, transparency, and replicability (epistemic values) (2023, 95–6). There is no explicit suggestion that epistemic values have precedence over non-epistemic values. In fact, both types of values are treated as starting points from where to think of best practices—both socially and empirically adequate—for the field of human genomics. But how, exactly, should these values be integrated into scientific practice?

We propose at least four ways in which values can play a legitimate role in human genomics:

1 *Values can help design research that is less potentially harmful*

In human genomics, current classifications based on ancestry groups tend to reflect old racial classifications, frequently dividing humans in terms of five continental populations (Eurasians/white; African/Black; Asians; Native Americans; and Oceanians). The proximity between these two classification schemes can easily result in slippage. An example of this can be seen in how the All of Us cohort study chose to use population descriptors such as ancestry and race interchangeably in their latest UMAP figure (Bick et al. 2024; see Section 41.4). This potential misuse of population descriptors generated controversy, as scholars flagged out the dangers of such a slippage between ancestry and race. This slippage can reinforce pernicious ideas about intrinsic

racial differences (Kozlov 2024). One of the main criticisms against that cohort study is that the UMAP figure ignored the “guiding principles” (values) recommended by the National Academies of Sciences’ report (2023). Values such as justice and respect encourage scientists to keep race and ancestry apart, only using racial categories in a few specific research circumstances (e.g., to study how racism impacts health). The cohort study infringed these values and recommendations, possibly contributing to unintended harm. In this sense, non-epistemic values can legitimately guide scientists to design research in less harmful ways.

Scientists can produce unwanted harmful consequences at different research stages, and thus, non-epistemic values are important throughout. In human genomics, the choice of software tools for data clustering and visualization is particularly important. For instance, PCA, STRUCTURE, and UMAP are well-known tools involving different idealizations that can lead to misinterpretations and harmful data interpretations. The algorithm STRUCTURE tends to emphasize genetic discrete groups by finding patterns of genetic similarity that maximize discreteness. This maximization can easily distort phylogenetic relationships and overestimate the genetic difference among populations (Winther 2022). Moreover, the output of this software STRUCTURE is usually presented as color-coded bar plots where clusters of similarity are presented in different colors. Therefore, while individuals may be similar on one measure, differences are overemphasized through the use of different colors producing clear-cut distinct clusters. The resulting bar plots reinforce the idea of discreteness, when in fact genetic variation across the globe is continuous. In such cases, choices about data analysis and interpretation should also consider non-epistemic values. Some software tools might be empirically adequate but vary in their ethical consequences. If this is the case, it would be irresponsible if scientists disregard these consequences when choosing what software to adopt.

2 *Values can help to choose between alternative hypotheses and conceptualizations*

The underdetermination argument states that scientists must appeal to background assumptions to use data to confirm a hypothesis (Section 41.2). Data itself cannot confirm a hypothesis, and the same data can be compatible with different hypotheses (Elliott 2022). The consequence is that often scientists choose between hypotheses because they disagree on the background assumptions. This choice is informed by non-epistemic values. For example, consider two hypotheses: “Graham has Northwestern European genetic ancestry” and “Graham is genetically similar to the genetic reference panel GBR 1000 Genome sample (on its 10 first principal components)” (Coop 2023). These are very similar hypotheses. In fact, the first hypothesis is an extrapolation of the second, as it uses genetic similarity to infer genetic ancestry relative to a data bank (1000 Genomes). This first hypothesis relies on assumptions about the relation between similarity and ancestry that the second hypothesis does not. Thus, the first hypothesis might have non-epistemic consequences that the second one does not, such as the suggestion that Graham belongs to a natural group and has a specific identity. These suggestions might conflict with how Graham understands his own identity. While data does not help us to choose between the two hypotheses, those non-epistemic consequences can. Recently, scientists proposed the conceptual shift from genetic ancestry to genetic similarity precisely because they understand that the latter is a more transparent, inclusive, and ethically harmless conceptualization (Coop 2023; National Academies of Sciences 2023). These non-epistemic values can help scientists to choose between alternative conclusions and their underlying concepts.

3 *Values can be introduced to manage inductive risk*

Accepting or rejecting a hypothesis are inductive activities and, as such, not immune to error. Scientists can wrongly accept a false hypothesis (false positive) or wrongly reject a correct one (false negative). Insofar as these mistakes might have foreseeable non-epistemic consequences, socially responsible scientists should use non-epistemic values when deciding what counts as sufficient evidence to support a hypothesis. Havstad’s (2022) discussion of genetic studies is a

relevant example. Studies demonstrated that Papua New Guineans have significant Denisovan ancestry in comparison to other non-African populations (Krause et al. 2010; Reich et al. 2010). However, the evidence for this claim comes from genetic material extracted from a tiny bone, and several scientists have rightly flagged the evidence as insufficient. The chance of error in those studies is high because it is based on only one high-coverage nuclear genome (2022, 310). As Havstad shows, those studies had non-epistemic consequences, such as reinforcing racial stereotypes of current Papua Guineans as a “primitive” population (2022, 304). The argument of inductive risk states that non-epistemic values such as anti-racism have a role to play in this context. Scientists would recognize the harm caused by racism, and thus, they would be particularly careful before concluding that Papua New Guineans have significant Denisovan ancestry. This would require raising evidential standards and collecting more evidence about that ancestry. In this case, anti-racism plays an indirect role because it helps scientists to decide how much evidence is sufficient for a claim and how much uncertainty is tolerable when accepting or rejecting a hypothesis.

4 **Values can guide the choice of research goals to be pursued**

Different non-epistemic values may shape different research goals, raising the question of what goals *should* human genomics pursue. One strategy is to consider which values help promote democratically endorsed epistemological and social aims of research (Intemann 2015; Schroeder 2021). Such aims of research are particularly important in the field of human genomics since it is a field known to have excluded and under-represented various human sub-groups by overrepresenting white individuals of European ancestry in most large-scale studies up to now (Hindorff et al. 2018; Popejoy and Fullerton 2016). Overrepresentation can be mitigated by producing science according to norms and standards that will promote democratic goals, such as equal representation of individuals from diverse ethnicities in studies. This has important downstream consequences since results stemming from human genomics provide us with a better understanding of the genetic origin of diseases, potentially carrying downstream consequences for different human groups. Some recent initiatives in the human genomics community seem to be taking a step in the right direction. For instance, the international consortium H3Africa (2012-) involved more than 32 countries and funding agencies such as Wellcome Trust (UK) and NIH (USA). This project aims at cataloguing a representative sample of African genetic diversity, but also to serve as a training platform and technology transfer to boost African human genomics for the benefit of Africans. As this example illustrates, democratic goals can legitimately shape large human genomic projects for the benefit of diverse populations across the globe.

An exhaustive list is beyond the scope of this chapter. Moreover, the list above is not exclusive. One and the same example may fall under many of the four categories above. Thus, the list of ways in which values can legitimately influence human genomics illustrates that non-epistemic values can be beneficial in different ways and that many geneticists already recognize such benefits explicitly or not. In the final section, we alert for the risks of failing to fully consider the role of non-epistemic values in human genomics.

41.4 A case study: *All of Us*

A concrete example where classification plays a central role in human genomics can be seen in the *All of Us* cohort study, launched in 2018 by the National Institutes of Health. Ultimately, the results and comparisons obtained will provide insights to develop tailored treatments through the growing field of precision medicine. Participants’ genomes are sequenced and grouped by genetic similarity on specific loci of their genomes by assessing single-nucleotide polymorphisms (SNPs). Significantly greater in scale than previous studies, *All of Us* aims to collect 1 million diverse genomes across

ethnicities (such as Hispanic or Latino), races (such as Black or African American and white) and ancestry groups (such as African and East Asian).

The most recent publication from the *All of Us* program gave rise to many concerns about the use of race, ethnicity, and ancestry (Bick et al. 2024). The study's latest UMAP figure 2 indicates a salient example of slippage (Bick et al. 2024, 342). First, the categories indicating ethnicity (such as Hispanic/Latino) and race (such as Black/African American) are combined in a representation, wrongly suggesting consistency or commensurability among these categories. However, the membership criteria for ethnicity and race are very different, making a comparison between them meaningless. This conflation illustrates how easily the slippage between those labels is even for highly trained geneticists. Second, scientists use matching colors to represent races, ethnicities, and ancestry groups, giving the misleading illusion that these categories are equivalent. For example, yellow represents both the race "white" and the ancestry group "European" across the figure. The visual suggestion is that Europeans correspond to the white race and vice-versa, but this is notably false. The white race is taken to be a socially constructed group based on self-reporting and it does not entirely overlap with the ancestry group of people with most European ancestry. This slippage entails an ethical risk because it facilitates misinterpretations and the perpetuation of dangerous views, such as that races have biological basis and that can effectively harm science and society. Slippage could be prevented or alleviated if scientists were more attuned to non-epistemic values like anti-racism and adjusted their representations of data (such as in the UMAP figure) accordingly.

At this point, one might wonder about the actual values that informed the use of population descriptors by the *All of Us* study and whether they align with the overarching goals of that study. If there is such an alignment, one might argue that using "race," "ethnicity," and "ancestry" is legitimate despite the problems with the UMAP figure. Unfortunately, the study and research program provides little information about those values and choices. First, the *All of Us* program website lists core values such as transparency, trust, and openness that guide the program as a whole, but it is unclear how those values apply to the specific case of choosing population descriptors. Second, the most recent study provides information related to "ancestry," but does not explain why comparisons with "race" and "ethnicity" were made. These factors illustrate that it is hard to assess the legitimacy of non-epistemic values by considering the goals of research. Frequently, the goals motivating scientific choices are opaque and are not made explicit. In the case of the *All of Us* program, we are left with some speculation. The use of "ancestry," "race," and "ethnicity" seems driven by goals of cross-study consistency and facilitating data use (Bick et al. 2024, "Methods" section). For instance, by using those different population descriptors, *All of Us* increases the chance of the data being used by scientists relying on any of those descriptors. Facilitating data use might be a good reason (value) to use those population descriptors. However, it is important to ask if this value trumps the potential harm caused by the slippage, the reification of racial differences, and other ethical risks described previously. One might wonder whether these potential harms justify a conceptual shift from genetic ancestry to genetic similarity, following the recommendations of the National Academies of Sciences (2023). One of the main issues with the *All of Us* program is that it has not been following these recommendations.

One important social value for science today is diversity, and this value seems also present in the *All of Us* program. Valuing diversity means that members of historically underrepresented groups in genomics research are adequately included, yielding promising results that may benefit such groups. In this sense, the program exemplifies how values can influence research goals. Lack of diversity in genomics is undesirable insofar as often results cannot extrapolate to different groups (Henrich, Heine, and Norenzayan 2010; Kidd et al. 2012; Martin et al. 2017). According to the most recent *All of Us* publication, it aims to increase genetic data from under-represented groups. 45% of its data comes from individuals who self-identify as non-European races or ethnicities (Bick et al. 2024, 345). The *All of Us* program also encourages diversity in the sense of including participants in decisions

regarding research priorities, information privacy, and other aspects of the program. It seems desirable to value diversity (in both senses above) insofar as it helps mitigate biases in genomics that have important downstream consequences for other areas such as public health and the development of drugs and treatments. However, more transparency would be required from the program to understand how diversity plays out in practice. It is important to consider what types of decisions participants are excluded from and to what extent this exclusion promotes power imbalances and pernicious value biases. The very fact that the UMAP picture was published raises the question about who took part in that decision and, most importantly, who did not.

41.5 Conclusion

In this chapter, we have provided an overview of the role of values in human genomics. We started by assessing whether value-freedom is desirable for genomics and highlighted some of the problems with this ideal. We have established that genomics is value-laden and should remain so. Then, we have identified four ways in which values can legitimately influence human genomics, with a focus on the use of population descriptors such as “race,” “ethnicity” and “ancestry.” Finally, we discussed the concrete example of the All of Us program and how problems might arise given the lack of attention to the influence of values in that field.

Understanding the role of values in human genomics is crucial. Scientists themselves recognize the role of values throughout different stages of genomics research (National Academies of Science, Engineering and Medicine 2023). This recognition is instrumental for a more reflective development of science in which geneticists are fully aware and capable of making informed decisions that impact society. Philosophical analysis of values in science (as the ones described in this chapter) can be helpful to this end. However, it is yet not clear whether recent approaches can offer feasible and actionable frameworks for scientific decision-making (Elliott and McKaughan 2014; Intemann 2015; Schroeder 2021). For that, scientists, philosophers, policymakers, and stakeholders have to work together.

Acknowledgments

Celso Neto was funded by ERC-STG-2024 HUGERA—Grant number 101161994, and Yasmin Haddad was funded by the Fonds de Recherche du Québec—Société et Culture—B3Z Postdoctoral Fellowship. The authors thank the volume editors for thoughtful comments and suggestions on earlier drafts of this chapter.

Note

- 1 The distinction between epistemic and non-epistemic values is not always clear and has generated debate among philosophers of science. Here we adopt this distinction for heuristic reasons and to follow standard terminology found in the literature, while acknowledging that in practice—and this is especially clear in the field of genomics—these values are often intertwined and inseparable.

References

- Ahn, Soohyun. 2020. “How Non-Epistemic Values Can Be Epistemically Beneficial in Scientific Classification.” *Studies in History and Philosophy of Science Part A* 84: 57–65. <https://doi.org/10.1016/j.shpsa.2020.08.002>.
- Anderson, Elizabeth. 2004. “Uses of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce.” *Hypatia* 19 (1): 1–24. <https://doi.org/10.1111/j.1527-2001.2004.tb01266.x>.
- Bick, Alexander G., Ginger A. Metcalf, Kelsey R. Mayo, et al. 2024. “Genomic Data in the All of Us Research Program.” *Nature* 627: 340–6. <https://doi.org/10.1038/s41586-023-06957-x>.

- Biddle, Justin. 2013. "State of the Field: Transient Underdetermination and Values in Science." *Studies in History and Philosophy of Science Part A* 44 (1): 124–33. <https://doi.org/10.1016/j.shpsa.2012.09.003>.
- Biddle, Justin B., and Rebecca Kukla. 2017. "The Geography of Epistemic Risk." In *Exploring Inductive Risk: Case Studies of Values in Science*, edited by Kevin C. Elliott and Ted Richards. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190467715.003.0011>.
- Bolnick, Deborah A. 2008. "Individual Ancestry Inference and the Reification of Race as a Biological Phenomenon." In *Revisiting Race in a Genomic Age*, edited by Barbara A. Koenig, Sandra Soo-Jin Lee, and Sarah S. Richardson. Rutgers University Press.
- Borrell, Luisa N., Jennifer R. Elhawary, Elena Fuentes-Afflick, et al. 2021. "Race and Genetic Ancestry in Medicine—A Time for Reckoning with Racism." *New England Journal of Medicine* 384 (5): 474–80. <https://doi.org/10.1056/NEJMms2029562>.
- Brown, Matthew J. 2013. "Values in Science beyond Underdetermination and Inductive Risk." *Philosophy of Science* 80 (5): 829–39. <https://doi.org/10.1086/673720>.
- Byeon, Yen Ji Julia, Rezarta Islamaj, Lana Yeganova, et al. 2021. "Evolving Use of Ancestry, Ethnicity, and Race in Genetics Research—A Survey Spanning Seven Decades." *The American Journal of Human Genetics* 108 (12): 2215–23. <https://doi.org/10.1016/j.ajhg.2021.10.008>.
- Coop, Graham. 2023. "Genetic Similarity versus Genetic Ancestry Groups as Sample Descriptors in Human Genetics." arXiv:2207.11595. Preprint, arXiv, January 7. <http://arxiv.org/abs/2207.11595>.
- Dauda, Bege, Santiago J. Molina, Danielle S. Allen, et al. 2023. "Ancestry: How Researchers Use It and What They Mean by It." *Frontiers in Genetics* 14: 1044555. <https://doi.org/10.3389/fgene.2023.1044555>.
- Douglas, Heather E. 2003. "The Moral Responsibilities of Scientists (Tensions between Autonomy and Responsibility)." *American Philosophical Quarterly* 40 (1): 59–68. <https://www.jstor.org/stable/20010097>.
- Douglas, Heather E. 2009. *Science, Policy and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin C., and Daniel J. McKaughan. 2014. "Nonepistemic Values and the Multiple Goals of Science." *Philosophy of Science* 81 (1): 1–21. <https://doi.org/10.1086/674345>.
- Elliott, Kevin C., and Rebecca Korf. 2024. "Values in Science: What Are Values, Anyway?" *European Journal for Philosophy of Science* 14 (4): 53. <https://doi.org/10.1007/s13194-024-00615-3>.
- Gannett, Lisa. 2004. "The Biological Reification of Race." *The British Journal for the Philosophy of Science* 55 (2): 323–45. <https://doi.org/10.1093/bjps/55.2.323>.
- Gannett, Lisa. 2014. "Biogeographical Ancestry and Race." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 47 (Part A): 173–84. <https://doi.org/10.1016/j.shpsc.2014.05.017>.
- Gould, Stephen Jay. 1993. *The Mismeasure of Man*. Norton.
- Haddad, Yasmin. 2025. "The Value-Ladenness of Ancestry." *Studies in History and Philosophy of Science* 112: 23–32. <https://doi.org/10.1016/j.shpsa.2025.06.001>.
- Harden, Kathryn Paige. 2021. *The Genetic Lottery: Why DNA Matters for Social Equality*. The Genetic Lottery. Princeton University Press. <https://doi.org/10.1515/9780691226705>.
- Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "Most People Are Not WEIRD." *Nature* 466 (7302): 29. <https://doi.org/10.1038/466029a>.
- Hindorff, Lucia A., Vence L. Bonham, Lawrence C. Brody, et al. 2018. "Prioritizing Diversity in Human Genomics Research." *Nature Reviews Genetics* 19 (3): 175–85. <https://doi.org/10.1038/nrg.2017.89>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- Kampourakis, Kostas, and Erik L. Peterson. 2023. "The Racist Origins, Racialist Connotations, and Purity Assumptions of the Concept of 'Admixture' in Human Evolutionary Genetics." *Genetics* 223 (3): iyad002. <https://doi.org/10.1093/genetics/iyad002>.
- Kidd, Jeffrey M., Simon Gravel, Jake Byrnes, et al. 2012. "Population Genetic Inference from Personal Genome Data: Impact of Ancestry and Admixture on Human Genomic Variation." *The American Journal of Human Genetics* 91 (4): 660–71. <https://doi.org/10.1016/j.ajhg.2012.08.025>.
- Kozlov, Max. 2024. "'All of Us' Genetics Chart Stirs Unease over Controversial Depiction of Race." *Nature News*, February 23. <https://doi.org/10.1038/d41586-024-00568-w>.

- Krause, Johannes, Qiaomei Fu, Jeffrey M. Good, et al. 2010. "The Complete Mitochondrial DNA Genome of an Unknown Hominin from Southern Siberia." *Nature* 464 (7290): 894–7. <https://doi.org/10.1038/nature08976>.
- Lewis, Anna C. F., Santiago J. Molina, Paul S. Appelbaum, et al. 2022. "Getting Genetic Ancestry Right for Science and Society." *Science* 376 (6590): 2502. <https://doi.org/10.1126/science.abm7530>.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Science as Social Knowledge. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Longino, Helen E. 2002. *The Fate of Knowledge*. Princeton University Press.
- Longino, Helen E. 2019. *Studying Human Behavior: How Scientists Investigate Aggression and Sexuality*. University of Chicago Press.
- Ludwig, David. 2016. "Ontological Choices and the Value-Free Ideal." *Erkenntnis* 81 (6): 1253–72. <https://doi.org/10.1007/s10670-015-9793-3>.
- Martin, Alicia R., Christopher R. Gignoux, Raymond K. Walters, et al. 2017. "Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations." *The American Journal of Human Genetics* 100 (4): 635–49. <https://doi.org/10.1016/j.ajhg.2017.03.004>.
- National Academies of Science, Engineering and Medicine. 2023. *Using Population Descriptors in Genetics and Genomics Research: A New Framework for an Evolving Field*. National Academies Press. <https://doi.org/10.17226/26902>.
- National Institute of Health. 2020. "All of Us Research Program." All of Us Research Program | NIH. January 6, 2020. <https://allofus.nih.gov>. Accessed July 7, 2025.
- Novembre, John. 2016. "Pritchard, Stephens, and Donnelly on Population Structure." *Genetics* 204 (2): 391–3. <https://doi.org/10.1534/genetics.116.195164>.
- Panofsky, Aaron, and Catherine Bliss. 2017. "Ambiguity and Scientific Authority: Population Classification in Genomic Science." *American Sociological Review* 82 (1): 59–87. <https://doi.org/10.1177/0003122416685812>.
- Popejoy, Alice B. 2021. "Too Many Scientists Still Say Caucasian." *Nature* 596 (7873): 463. <https://doi.org/10.1038/d41586-021-02288-x>.
- Popejoy, Alice B., and Stephanie M. Fullerton. 2016. "Genomics Is Failing on Diversity." *Nature* 538 (7624): 161–4. <https://doi.org/10.1038/538161a>.
- Reardon, Jenny. 2005. *Race to the Finish: Identity and Governance in an Age of Genomics*. Princeton University Press. <https://www.jstor.org/stable/j.ctt7t00f>.
- Reich, David, Richard E. Green, Martin Kircher, et al. 2010. "Genetic History of an Archaic Hominin Group from Denisova Cave in Siberia." *Nature* 468 (7327): 1053–60. <https://doi.org/10.1038/nature09710>.
- Reydon, Thomas A. C., and Marc Ereshefsky. 2022. "How to Incorporate Non-Epistemic Values into a Theory of Classification." *European Journal for Philosophy of Science* 12 (1): 4. <https://doi.org/10.1007/s13194-021-00438-6>.
- Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber et al. 2002. "Genetic Structure of Human Populations." *Science* 298 (5602): 2381–5. <https://doi.org/10.1126/science.1078311>.
- Schroeder, S. Andrew. 2021. "Democratic Values: A Better Foundation for Public Trust in Science." *The British Journal for the Philosophy of Science* 72 (2): 545–62. <https://doi.org/10.1093/bjps/axz023>.
- Spencer, Quayshawn. 2014. "A Radical Solution to the Race Problem." *Philosophy of Science* 81 (5): 1025–38. <https://doi.org/10.1086/677694>.
- Uffelmann, Emil, Qin Qin Huang, Nchangwi Syntia Munung, et al. 2021. "Genome-Wide Association Studies." *Nature Reviews Methods Primers* 1 (1): 1–21. <https://doi.org/10.1038/s43586-021-00056-9>.
- Winther, Rasmus Grønfeldt. 2022. *Our Genes: A Philosophical Perspective on Human Evolutionary Genomics*. Cambridge University Press. <https://doi.org/10.1017/9781316756324>.
- Wonkam, Ambroise. 2021. "Sequence Three Million Genomes across Africa." *Nature* 590 (7845): 209–11. <https://doi.org/10.1038/d41586-021-00313-7>.
- Yudell, Michael. 2011. "A Short History of the Race Concept." In *Race And the Genetic Revolution: Science, Myth and Culture*, edited by Sheldon Krimsky and Kathleen Sloan. Columbia University Press.

PHILOSOPHICAL APPROACHES TO VALUES IN CLIMATE SCIENCE

Greg Lusk

42.1 Introduction

Given its close association with the pressing social issue of climate change, climate science has long wrestled with accusations that a social agenda lurks behind its investigations. This agenda purportedly aligns with broadly leftist or liberal interests. For example, right-leaning commentators have accused those concerned with climate change of being political watermelons: green on the outside and red on the inside (e.g., [Delingpole 2011](#); also see [Oreskes and Conway 2010](#)). Climate scientists have also been accused of fraud. For example, Phil Jones, then director of the Climatic Research Unit at the University of East Anglia, was accused of using inappropriate data analysis techniques to help ensure warming was seen in his climate analysis. The accusation was based on comments in his email messages that were hacked in the 2009 “Climategate” affair. The comments turned out to be misunderstood, and after much scrutiny, he and his colleagues were cleared of any wrongdoing ([McKie 2019](#)). More recently, Patrick Brown, co-director of the Climate and Energy Team at the Breakthrough Institute, claimed that a 2023 paper he published in the prestigious journal *Nature* had to push a climate change narrative to be accepted for publication (see [Brown 2023](#); [Brown et al. 2023](#)). In response, the journal released the peer review reports for the paper, which seemed to cast doubt on Brown’s narrative ([Ward 2023](#)).

Perhaps such accusations are unsurprising. Evidence suggests that scientists of most stripes lean politically left ([Kaurov et al. 2022](#)) and critics of climate action often come from the political right. What may seem more surprising, however, is that social values *do* play a significant role in climate science. Their role is often more nuanced, and often less threatening, than the claims of bias that make news headlines. Nonetheless, their role is significant enough that caution and care are needed. More attention should be paid to when and where social values influence climate science.

To that end, this chapter has a dual focus. It first examines some of the crucial locations where social values play a role in climate science, and second, it analyzes different positions on managing that value influence. The chapter also has a distinct philosophical perspective: it highlights value influences and responses that have become the touchstones for conversations in philosophy of science.

Before beginning, it is helpful to discuss “values.” There is often a distinction drawn in philosophy of science between epistemic and non-epistemic values. Exactly what that distinction should be, however, is contentious, and many of the authors discussed below use different notions. Those concerned with such a distinction should read Elliott and Richards’s introduction to this *Handbook*

(Elliott and Richards 2026). I'm going to assume that epistemic values are indicators of truth, or some other cognitive good, and only truth or only another cognitive good. I talk very little about epistemic values in this chapter. I'll use the terms non-epistemic values and social values, or just unqualified "values," to talk about anything that is non-epistemic. Non-epistemic or social values can be values that have a political, personal, or religious nature (e.g., value of human life, desire for profit, etc.), but are not indicators of any epistemic good. This umbrella definition smooths over some small differences in the positions I discuss in favor of a more cohesive narrative.

42.2 Inductive risk and climate science

Like with many other areas of science, philosophical interest in the role of social values in climate science dovetailed with a renewed interest in arguments demonstrating that social values had an ineliminable, and perhaps even desirable, role in scientific reasoning. The most influential of these arguments, now referred to as the argument from inductive risk, was originally presented by Richard Rudner in the 1950s (see Rudner 1953) but was then brought to prominence and expanded upon by Heather Douglas (2000, 2009). Interested readers should refer to Elliott and Richards (2026), Elliott (2026), or Brown (2026) in this *Handbook* for more details on the value-free ideal, its demise, and inductive risk. As those chapters indicate, the crux of inductive risk arguments is the existence of epistemically unforced decisions; it is in making such decisions that non-epistemic values enter the core of scientific practice.

It is worth looking at some examples to see the myriad of epistemically unforced decisions that one might face within the context of climate science. A notable example concerns at what point one should accept the hypothesis that anthropogenic factors are influencing the climate system, and how forcefully to deliver that message (see Elliott 2017, chap. 5). In 1988, the then Director of NASA's Goddard Institute of Space Studies, James Hansen, testified in front of the US Congress that "Global warming has reached a level such that we can ascribe with a high degree of confidence a cause and effect relationship between the greenhouse effect and observed warming ... It is already happening now" (Shabecoff 1988). Perhaps more importantly Hansen reported high confidence that the warming was due to the release of greenhouse gases, and not random fluctuations.

Hansen's testimony is now looked at as a pivotal moment for the visibility of climate change as a serious threat to human well-being. His testimony occurred at an exceptionally warm period of time across most of the United States, and some credit this weather with bringing attention to Hansen's conclusions (he had testified in previous years, but his message received little attention). But why testify then, in the late 1980s? After all, he and his NASA colleagues could continue to study the climate looking for global warming and its consequences. And why ascribe a high confidence to the human influence on the climate?

When to accept a hypothesis, and how best to communicate results, is subject to inductive risk considerations of the kind mentioned above. There is always a risk of error—for example, that the global warming hypothesis was wrong—even though the evidence suggests otherwise. There is risk that the way one communicates results will result in the message being misheard or misunderstood. Here non-epistemic factors and social values come into play. This is precisely how Hansen justified his testimony and messaging; he claims to have weighed the *social* costs of being wrong versus the costs of not talking (Weart 2025). Essentially, he chose to announce that global warming had been detected because not bringing this scientific information to light might have *significant human costs*. The desire to prevent human costs is a social value.

Inductive risk arguments do not just apply to hypothesis acceptance and communication in climate science; there are many sites where epistemically unforced choices exist, and social values

must or should play a role. [Parker and Lusk \(2019\)](#) illustrate an inductive risk approach to climate-related decision-making using the selection of observation instruments. In their example, scientists must install one of two instruments, where neither is known to be better than the other; that is, the choice is epistemically unforced. One can reasonably then appeal to social values to make the selection. For example, if instrument A might be expected to underestimate a quantity, and instrument B overestimate it, one could reasonably select B if the quantity of interest might be harmful to human well-being. Instrument B is justified because it would be better for human well-being to overestimate dangers and then put in place protections, rather than underestimate and risk human harm. Making this choice is justified by the non-epistemic or social value of protecting human life.

The above pattern of reasoning can be applied in many different sites within climate science. [Parker and Lusk \(2019\)](#) further explore the way values might influence selecting information for analysis, choosing a data analysis technique, and estimating uncertainties. Kristen [Intemann \(2015\)](#) examines how values influence causal claims, and epistemic tradeoffs. Several individuals have examined how the analysis of extreme weather is impacted by social factors (e.g., [Donhauser 2021](#); [Intemann 2015](#); [Lusk 2017](#)). There is no shortage of unforced methodological choices in climate science. However, this is true across most scientific disciplines. Thus, it is tempting to see climate science as unexceptional from an inductive risk perspective: there are similar decisions made in lots of socially relevant science. Nevertheless, there are practices in climate science that are somewhat unique and also influenced by values, and it is to these that we now turn.

42.3 Non-epistemic values in global climate modeling

Computer models play a central role in climate science. Such models serve many purposes, including detecting the signal of human influence on the climate, making projections of future climate under various emissions scenarios, and analyzing the changed odds of extreme weather. These models come in the form of mathematical representations of the climate system that include characterizations of the atmosphere, ocean, land masses, sea ice, and other factors that influence climatic conditions. Creating a state-of-the-art global model is a massive undertaking requiring large teams of scientists, coders, hardware specialists, as well as a supercomputer to run the model. Given the significant investment and specialized hardware required, only dedicated research institutions and government entities (e.g., the National Center for Atmospheric Research in the United States) tend to develop, run, and maintain models.

Philosophers have found climate models interesting in part because they are not candidates for truth. Scientists are aware that their climate models contain many false and merely approximate elements, because they purposely included these elements in the model. For example, scientists often employ parameterizations—simplified representations of small-scale processes—to portray elements like convection and cloud formation, because including detailed representations would be too computationally costly. Thus, truth is not often the central concern when assessing the quality of a climate model.

Instead, climate models are often assessed through the examination of their adequacy for purpose (see [Parker 2009, 2020](#)). On the adequacy-for-purpose approach, climate models are assessed relative to certain goals or purposes for which the model will be used. Instead of trying to confirm that a model is true, models would be “selected based on their ability to answer questions of interest” ([Taper, Staples, and Shepard 2008](#), 358) and assessing a model would involve “an activity that aims to determine whether a model is adequate for one or more purposes of interest” ([Parker 2010](#), 287; see also [Parker 2020](#)). These purposes and interests could be related to what those using the model are interested in, for example, temperature changes in a particular region. Or the purposes could relate to functions the model might serve, for example, a model might be adequate only if it could be reliably used within a particular regulatory regime. More than likely, it will involve some mix of both.

The establishment of purposes for model construction and evaluation is one way in which non-epistemic values influence the modeling process (see [Lusk and Elliott 2022](#)). Selecting a purpose is often a situation where one needs to select between different desirable characteristics. After all, we cannot expect one model to be adequate for all, or even many, of our purposes. A model that is adequate for the purposes of examining drought in North America might not be adequate for examining floods in Asia, choosing locations of wind turbines in Europe, or for the purpose of modeling convective storms (like hurricanes). Which purposes a model is developed for or is assessed against reflect non-epistemic values that reveal a set of preferences or perceived degrees of importance.

Eric Winsberg has argued that developing and assessing models with regard to their value-laden purposes results in a kind of “prediction preference” that can be hard to eliminate. As Winsberg notes “When it comes to model construction, for instance, purposes and priorities shape not just which entities and processes are represented in a model, but also *how* they are represented, including which simplifications, idealizations, and other distortions are more or less tolerable” ([Winsberg 2018](#), 139). However, as we noted earlier, it is a difficult task to build a global climate model, and thus, scientists typically choose to use and adapt *existing models* rather than rebuild from scratch.

The reuse of models for different purposes creates a chain of value influence that can be hard to undo. After all, if a model is built for certain purposes at certain time but then is reused at a later time for different purposes, the model still contains all the representative elements that were selected to support the initial purpose. It is difficult (and often infeasible) to change those elements later to make them better fit for a different purpose (see [Winsberg 2018](#), 141–4). After all, it may not be obvious when reusing a model which choices were made for what reasons, and many elements are usually balanced against each other to ensure adequacy-for-purpose. The original purposes thus create a kind of path dependency for the model, once it has been designed and assessed for those purposes, unless overhauled, future results will bear the mark of choices aligned with a particular purpose. Given the close connection between values and purposes, it would be difficult if not impossible to go back and scrub all the value influence from a model.

42.4 Values in integrated assessment models

Integrated assessment models (IAMs) differ in kind from the global climate models discussed above in that they include an economic component to investigate how changes in the climate system will impact economic growth. Such models can be used to analyze the consequences of policies—like climate mitigation measures—on both the economic and climate system. Or they might be used to compare different environmental policies by ranking them in terms of their preferred outcomes. IAMs can sometimes be built for specific economic sectors (like agriculture or energy), to see how particular groups will be impacted by policies. IAMs are often seen as especially useful tools for policy under a cost–benefit style of assessment, where a climate-related action would be recommended when the future benefit of the action is equal to or exceeds the action’s present (and continuing) costs.

IAMs are typically used to compute an important number for cost–benefit analysis: the social cost of carbon (SCC). The SCC is a financial estimate of the damage of adding one additional ton of CO₂ to the atmosphere. If the cost of a climate abatement policy is less than the SCC produced in the absence of the policy (i.e., the abatement is cheaper), then cost–benefit analysis should dictate adopting the policy. The SCC is most prominently employed by the United States Federal Government, several individual state governments, as well as the Government of Canada. Some call the SCC “the single most important number for thinking about climate change” (see Marshal Burke in [Backman 2021](#)) and William Nordhaus won the Nobel Memorial Prize in Economic Sciences for his work on climate change economics, a central part of which was the development of an IAM, called the DICE model, and calculations of the SCC.

IAMs typically contain at least three components: (1) a simplified climate model (or other means of ingesting physical information) that specifies how the climate system will react to additional carbon emissions, (2) a “damage function” that tracks how economic activity will be altered in response to climate change, (3) a means of expressing and aggregating welfare across time in terms of a particular quantity (e.g., US dollars). Together, these components link environmental changes to the way economies function and the ways future humans (and as a consequence, many non-humans) might live.

Practitioners and critics both acknowledge that there are substantial uncertainties that invite value influence when designing and employing IAMs. As Beck and Kruger explain: “the existence of scientific and ethical uncertainty requires modelers to make numerous choices in model development: choices about model scope, equations, parameter values, and output presentation. These modeling decisions have epistemic, ethical, and political dimensions” (2016, 628).

Historically, one of these decisions is how to calibrate the simple climate component of the IAM to ensure its realistic behavior. One element involved in this calibration is selecting parameters for quantities like climate sensitivity, which tracks how the climate system responds to changes in atmospheric CO₂. As Mathias Frisch (2013) has pointed out, there are lots of different estimates of climate sensitivity provided by more complex global climate models, and no agreed-upon way to select between them or combine them for calibration. Frisch judges that a particular strategy (that had been previously employed in practice) “dangerously ignores the uncertainties in climate predictions and downplays possible risks” (122). Implicit in his argument is that one should take seriously the possible risks and select strategies that would make them more salient, which practitioners have since done. It seems clear that decisions like this are open to inductive risk considerations and value influence.

The influence of values extends to the economic components of an IAM as well. For example, a similar situation to the above arises when selecting a damage function: there are different forms of the damage function, no universally agreed-upon best function, and the chosen form will impact the results and could steer eventual decision-making.

A particularly high-profile example of values at work in integrated assessment modeling comes from the choice of social discount rate, which is used in computations of the SCC. Though we’ve already seen that some believe the SCC to be the most important number for thinking about climate change, what that number should be is contested, often because of disagreements about the discount rate. The discount rate controls the view of future costs and benefits. As J. Paul Kelleher explains it: “A discount rate close to zero treats a dollar’s worth of damage in the future as roughly on par with a dollar’s worth of damage today, whereas a discount rate of 10% treats a dollar’s worth of damage in 100 years as being much less costly than one cent’s worth of damage today” (Kelleher 2017, 958). Kelleher describes how influential this number can be: “Within any such analysis of climate policy, a small change in the value of a single technical parameter—the so-called social discount rate—can make the difference between recommending a go-slow approach to mitigating greenhouse gas emissions and recommending immediate and costly actions to curb them” (Kelleher 2017, 958).

While the selection of the discount rate is clearly influential from a social policy perspective, philosophers believe it is also a deeply ethical choice. Many philosophers follow what the IPCC Second Assessment report called a “prescriptivist approach” to social discounting. On this view, the social discount rate should be decided primarily on ethical grounds; that is, we should set the rate in accordance with the degree to which the future is valued relative to the present. Many think that future persons are ethically equivalent to present ones; discounting is thus an intergenerational justice concern. Such a view is associated with Nicholas Stern, lead author of the Stern Review on the Economics of Climate Change, commissioned by the UK Government and released in 2006. Stern writes “We take a simple approach in this Review: if a future generation will be present, we suppose that it has the same claim on our ethical attention as the current one” (Stern 2007, part 1, 31).

The prescriptivist approach is often contrasted with the descriptivist approach, followers of which, despite the name, still make climate policy recommendations but set their discounting rate using a different strategy. The descriptivist approach views discount rates as something like a reverse interest rate. The approach is closely associated with William Nordhaus, who writes, “As this approach relates to discounting, it requires that we look carefully at the returns on alternative investments—at the real interest rate—as the benchmarks for climatic investments. The normatively acceptable real interest rates prescribed by philosophers, economists, or the British government are irrelevant to determining the appropriate discount rate to use....” (Nordhaus 2007, 692). On this view, ethics shouldn’t enter the conversation about discount rates; rather, such rates can be determined by the history of markets and subsequent analysis.

What is very clear, however, is that these two approaches result in different costs for carbon and therefore may produce very different policy recommendations. The SCC that was obtained using Stern’s assumptions is ten times higher than the results from Nordhaus’ assumptions (Heal and Millner 2014). Costs considered reasonable for carbon ranged from \$30 to \$200 in 2021 (Mintz-Woo 2022). Such estimates have been politically contentious. Under the Biden administration, the US Environmental Protection Agency was using a cost of \$51 and in late 2023 proposed an increase to \$190 in future analyses (Hersher et al. 2023). These costs supersede those appealed to by the Trump Administration, which was estimated to be between \$1 and \$6 (Bordoff 2017).

It is worth noting that the choice of using IAMs and cost–benefit analysis is itself a value-laden decision, to which many object. It is difficult, and some would argue impossible, to place monetary value on environmental goods like clean air, water, and livable areas, and these goods may not be substitutable by others, including high incomes (see Gardiner 2011). Others have critiqued the short-term view of IAMs. Some of the CO₂ we emit now will remain in the atmosphere for thousands of years, applying any discount rate would render distant damages from these emissions invisible, and the projections that employ IAMs usually only extend for a hundred years or so at most. A group of geoscientists and a philosopher decided to calculate the cost of carbon at the geological scales on which the carbon cycle operates and estimated the “ultimate cost of carbon” to be \$100K, significantly above the standard price per ton (Archer, Kite, and Lusk 2020). It seems quite clear that assumptions motivated by different values can have a significant impact on the modeling that steers our response to climate change.

42.5 Dealing with values in climate science: Philosophical responses, strategies, and consequences

There is little doubt that values play a significant role in climate science, including its core “rational” elements. The obvious question that arises, then, is what should be done about the presence of those values.

Several philosophers have argued that values in climate science could be beneficial. For example, Parker and Lusk (2019) have argued that the inclusion of values within climate services (a particular area of climate science concerned with readying communities for climate change) can be used to enhance information uptake and ensure the usability of climate information. Essentially, Parker and Lusk argue that by embracing the values of those that use climate information (e.g., decision makers, municipalities, or other groups), scientists can tailor information to their needs and value systems. The resulting climate information will be more useful—the values employed within the science will align with the values of the users—thus promoting information uptake. The uptake of information will then help ensure that users avoid “unknowingly [making] decisions that fail to protect against the climate-related harms that they most want to avoid” (Parker and Lusk 2019, 1648).

Followers of what is sometimes called the “aims approach” to values in science also see values in climate science as acceptable, at least under certain conditions. As Kristen Intemann argues, values

in climate science are legitimate to employ “insofar as doing so will promote the democratically endorsed epistemological and social aims of research” (Intemann 2015, 218). For example, if there is a democratic preference for environmental safety, it might be legitimate to employ the worst-case scenario when making climate model projections. Lusk (2020; see also 2021) is similarly worried about politics and democratic legitimacy within the provision of climate services. They worry that, given how sensitive some climate information is to choices that involve values, the inclusion of values within climate science might allow scientists or others to covertly steer downstream policy through the backdoor via value selection. Still, they argue that appeals to values are legitimate so long as they are selected through deliberative democratic means. These democratic procedures help protect against scientists having an undue influence on decision-makers (for more on the connection between values, democracy and legitimacy see the chapter by Schroeder (Schroeder 2026) and the other chapters in Section 42.3 of this *Handbook*).

However, values may not be welcome in every nook of climate science, and embracing them is not the only management strategy. One can minimize the role of values in climate science by taming uncertainty and thus reducing the number of epistemically unforced decisions. Since reducing uncertainty is often beneficial in science, there are a number of strategies that climate scientists already employ. For example, climate scientists prefer to perform many model runs—producing what they call an *ensemble*—rather than just a single run to make future projections (see Frigg, Thompson, and Werndl 2015). In the ensemble, scientists may vary the parameters or initial conditions, or both. The hope is that if the value of something like a parameter is uncertain, scientists can at least explore a range of uncertainty by running the model with many different parameter values. While selecting the range of parameter values to include in an ensemble might still be a value-influenced choice, it would seemingly be less consequential than having to select only one parameter value and only report the consequences of that value, because one can see how the results depended on the values chosen. There are also multi-model ensembles that attempt to explore the uncertainties and differences in the different structure of models. Similarly, when making projections of the future, scientists often run their models under several different emissions trajectories, which reduces the need to select one future path of emissions. Of course there is a limitation to this strategy: exploring uncertainties uses significant computational and financial resources. Thus, making decisions about when to use an ensemble and how large it should be are likely to be value-laden in the inductive risk sense. Still, results that explore uncertainty make salient how results depend on choices made in the modeling procedure and thus give users of that information a way to track or account for those choices.

A similar strategy is used to reduce the influence of values on the way climate projections are reported, even when uncertainty is explored. As Wendy Parker (2014) points out, the results of climate model projections are artificially precise: even though results look relatively exact, scientists do not usually understand the climate system well enough to give a tightly constrained probability estimate for a detailed climate hypothesis. Rather than give precise probability estimates, scientists often report results using coarse language. Parker gives the example of the IPCC’s calibrated language from the 5th Assessment Report for conveying uncertainties that accompany certain results. For instance, the IPCC says that “likely” should accompany results that would have a probability assignment of 66–100%. Parker concludes that, “even if social values sometimes do come into play in the model development process...the influence of those values on estimates of uncertainty will be reduced when coarser estimates are given. The influence will be reduced insofar as choices in model development will less often *make a difference* to the uncertainty estimates produced” (Parker 2014, 28).

Some philosophers claim that the management of uncertainty can be used to eliminate value influence altogether. Gregor Betz (2013) attempts to use a variant of coarse estimate strategy to manage value influence in science, using climate science as his example. Betz argues that “Allegedly arbitrary and value-laden decisions can be systematically avoided...by making uncertainties explicit and

articulating findings carefully” (2013, 209). Betz claims that inductive risk proponents need more than just the existence of unavoidable epistemically unforced choices with social consequences to claim values are ineliminable in science; they need to show that the reports of subsequent scientific results carry “substantial error probabilities.” The way to eliminate those substantial error probabilities is to *hedge* claims. Hedging a claim is weakening it beyond a reasonable doubt such that its status is akin to plain facts like “the atmosphere comprises oxygen, coal burns, CO₂ is a greenhouse gas” (Betz 2013, 215). Betz claims that the IPCC is an example of value-free science. The IPCC uncertainty guidances, including the calibrated language (e.g., the “likely” label and probability interval) Parker appealed to above, are sufficient hedging to put claims beyond a reasonable doubt.

While often cited, Betz’s argument for eliminating value influence is unconvincing: hedging a complex claim about a future climate state so that it is beyond a reasonable doubt, but still informative for policy makers, seems difficult (e.g., it is hard to build policy on “coal burns”). Stephen John (2015) provides a more nuanced critique. He argues that hedged claims are formed relative to some body of evidence. For Betz’s hedging to be successful, a hedged claim must be formulated from a body of evidence that was created without error tradeoffs and decisions subject to inductive risk. If social values influenced the construction of the body of evidence, even hedged hypotheses based on that evidence will be influenced by the same values; the value influence just moves up a level. Insofar as the IPCC (roughly) uses peer review as its criteria for inclusion in the body of evidence it considers, the “IPCC’s construction of its body of evidence does involve making a significant judgment about how to trade-off risks of adopting false evidence claims versus failing to adopt true evidence claims” (John 2015, 6). He illustrates how this works in practice using studies of the West Antarctic ice sheet. Thus, hedging claims formed from curated evidence would not result in claims free from the error probabilities with which Betz is concerned. This does not mean that hedging is completely irrelevant (see Ward and Creel 2024), but hedging does not seem capable of making climate science value-free.

John himself advocates for maintaining the spirit of value freedom within climate science by employing high, fixed, epistemic standards. This doesn’t keep climate science value-free, but it does help ensure the production of information that is beyond a reasonable doubt, and that these standards can be justified by an *epistemic* preference for knowledge. In the case of the IPCC, though its work involves making decisions that trade-off inductive risks—for example when deciding how much to hedge and what information to include in its body of evidence—it displays a preference for avoiding false positives and obtaining knowledge.

Such a proposal may have significant negative consequences. Several scholars have suggested that climate science is already too conservative and should be more concerned with ensuring early warning to avoid the disastrous consequences of climate change than avoiding false positive results. Climate scientist James Hansen, for example, has complained that there is a danger in excessive caution when it comes to making claims about climate change. He suggests that we should rue this “scientific reticence” if it “serves to lock in future disasters” (Hansen 2007, 1). Brysse et al. (2013) have argued that climate scientists have tended to “err on the side of least drama” and are bias toward cautious estimates rather than alarmism (2013, 327). Similarly, Lloyd and Oreskes have argued that a preference among climate scientists to underestimate, rather than overestimate, the human impacts from climate change is one fact that led to a delay in accepting certain forms of extreme event attribution and analysis (Lloyd and Oreskes 2018). High fixed epistemic standards have consequences, potentially problematic ones.

Ahmad Elabbar (2023) claims that John’s proposal leads to climate injustice in some instances. Recall that the IPCC assessment process includes only high-quality peer-reviewed materials. Elabbar argues that this high and fixed standard produces an unequal distribution of epistemic power among stakeholders. We live in an evidentially unequal world and there is often less data where individuals tend to be less-well off (e.g., the global south). If assessment reports like the IPCC systematically

ignore data for certain regions because of poor quality, and no other data is available, then the rate of important findings will be smaller for those data-scarce regions, relative to data-rich environments. This has the potential to exacerbate climate injustice, as such findings are needed to argue for climate aid or to bring attention to climate problems. Epistemic power would be less unequal, Elabbar argues, if the standards were relaxed to include information about data-poor regions, even if that data fails to meet the normal IPCC standards. Thus, a concern for justice justifies employing varying, rather than high and fixed, epistemic standards.

One question that arises in the above debates is what exactly the immediate goal of climate science is (or should be). Intemann (2015) seems to indicate that the aim is something like environmental protection or human health; John seems to think it could be knowledge production; Elabbar suggests justice is a goal. Lusk and Elliott (2024) have recently argued that local aims for science (e.g., the protection of human or environmental health) are more defensible than global aims for science (e.g., knowledge or truth). More work needs to be done to tease out the role of aims and goals in this literature, but to some degree, the above debates suggest that such matters influence approaches to values in science, including climate science.

42.6 Conclusion

This chapter examined the role of social values within the philosophy of climate science. The first section examined sites where philosophers see substantial influence of values on climate science, with a focus on inductive risk and unforced decisions within climate and integrated assessment modeling. The second section examined responses to the presence of social values within climate science; while some thought values could beneficially promote the uptake of information or climate justice, others tried to limit the role of values in efforts to promote knowledge or political authority. How climate science can best play its role as an important producer of decision-relevant information while also producing high-quality knowledge is a question that remains open and deserves further attention.

References

- Archer, David, Edwin Kite, and Greg Lusk. 2020. "The Ultimate Cost of Carbon." *Climatic Change* 162 (4): 2069–86. <https://doi.org/10.1007/s10584-020-02785-4>.
- Backman, Isabella. 2021. "Professors Explain the Social Cost of Carbon." *Stanford Report: Explainers*, June 7. <https://news.stanford.edu/stories/2021/06/professors-explain-social-cost-carbon>.
- Beck, Marisa, and Tobias Krueger. 2016. "The Epistemic, Ethical, and Political Dimensions of Uncertainty in Integrated Assessment Modeling." *WIREs Climate Change* 7 (5): 627–45. <https://doi.org/10.1002/wcc.415>.
- Betz, Gregor. 2013. "In Defence of the Value Free Ideal." *European Journal for Philosophy of Science* 3: 207–20. <https://doi.org/10.1007/s13194-012-0062-x>.
- Bordoff, Jason. 2017. "Trump vs. Obama on the Social Cost of Carbon—and Why It Matters." *Wall Street Journal*, November 15. <https://www.wsj.com/articles/trump-vs-obama-on-the-social-cost-of-carbonand-why-it-matters-1510769071>.
- Brown, Mathew J. 2026. "Recent Arguments for the Ideal of Value-Free Science." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Brown, Patrick. 2023. "I Left Out the Full Truth to Get My Climate Change Paper Published." *The Free Press*, September 5. <https://www.thefp.com/p/i-overhyped-climate-change-to-get-published>.
- Brown, Patrick T., Holt Hanley, Ankur Mahesh, et al. 2023. "Climate Warming Increases Extreme Daily Wildfire Growth Risk in California." *Nature* 621 (7980): 760–66. <https://doi.org/10.1038/s41586-023-06444-3>.
- Brysse, Keynyn, Naomi Oreskes, Jessica O'Reilly, and Michael Oppenheimer. 2013. "Climate Change Prediction: Erring on the Side of Least Drama?" *Global Environmental Change* 23 (1): 327–37. <https://doi.org/10.1016/j.gloenvcha.2012.10.008>.
- Delingpole, James. 2011. *Watermelons: The Green Movement's True Colors*. Publius Books.
- Donhauser, Justin. 2021. "How to Make Value-Driven Climate Science for Policy More Ethical." *Studies in History and Philosophy of Science Part A* 89: 31–40. <https://doi.org/10.1016/j.shpsa.2021.06.014>.

- Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67 (4): 559–79. <https://doi.org/10.1086/392855>.
- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elabbar, Ahmad. 2023. "Varying Evidential Standards as a Matter of Justice." *British Journal for the Philosophy of Science*, August. <https://doi.org/10.1086/727429>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Elliott, Kevin C. 2026. "Arguments Against the Value-Free Ideal." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Elliott, Kevin C., and Ted Richards. 2026. "Introduction." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Frigg, Roman, Erica Thompson, and Charlotte Werndl. 2015. "Philosophy of Climate Science Part II: Modelling Climate Change." *Philosophy Compass* 10 (12): 965–77. <https://doi.org/10.1111/phc3.12297>.
- Frisch, Mathias. 2013. "Modeling Climate Policies: A Critical Look at Integrated Assessment Models." *Philosophy and Technology* 26 (2): 117–37. <https://doi.org/10.1007/s13347-013-0099-6>.
- Gardiner, Stephen M. 2011. *A Perfect Moral Storm: The Ethical Tragedy of Climate Change*. Oxford University Press.
- Hansen, James E. 2007. "Scientific Reticence and Sea Level Rise." *Environmental Research Letters* 2 (2): 024002. <https://doi.org/10.1088/1748-9326/2/2/024002>.
- Heal, Geoffrey M., and Antony Millner. 2014. "Agreeing to Disagree on Climate Policy." *Proceedings of the National Academy of Sciences* 111 (10): 3695–98. <https://doi.org/10.1073/pnas.1315987111>.
- Hersher, Rebecca, Rebecca Ramirez, Aaron Scott, and Margaret Cirino. 2023. "EPA's Proposal to Raise the Cost of Carbon Is a Powerful Tool and Ethics Nightmare." *NPR*, February 9. <https://www.npr.org/2023/02/08/1155544378/epas-proposal-to-raise-the-cost-of-carbon-is-a-powerful-tool-and-ethics-nightmar>.
- Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5 (2): 217–32. <https://doi.org/10.1007/s13194-014-0105-6>.
- John, Stephen. 2015. "The Example of the IPCC Does Not Vindicate the Value Free Ideal: A Reply to Gregor Betz." *European Journal for Philosophy of Science* 5 (1): 1–13. <https://doi.org/10.1007/s13194-014-0095-4>.
- Kaurov, Alexander A., Viktoria Cologna, Charlie Tyson, and Naomi Oreskes. 2022. "Trends in American Scientists' Political Donations and Implications for Trust in Science." *Humanities and Social Sciences Communications* 9 (1): 1–8. <https://doi.org/10.1057/s41599-022-01382-3>.
- Kelleher, Paul J. 2017. "Descriptive Versus Prescriptive Discounting in Climate Change Policy Analysis." *Georgetown Journal of Law and Public Policy* 15: 957–77.
- Lloyd, Elisabeth A., and Naomi Oreskes. 2018. "Climate Change Attribution: When Is It Appropriate to Accept New Methods?" *Earth's Future* 6 (3): 311–25. <https://doi.org/10.1002/2017EF000665>.
- Lusk, Greg. 2017. "The Social Utility of Event Attribution: Liability, Adaptation, and Justice-Based Loss and Damage." *Climatic Change* 143 (1–2): 201–12. <https://doi.org/10.1007/s10584-017-1967-3>.
- Lusk, Greg. 2020. "Political Legitimacy in the Democratic View: The Case of Climate Services." *Philosophy of Science* 87 (5): 991–1002. <https://doi.org/10.1086/710803>.
- Lusk, Greg. 2021. "Does Democracy Require Value-Neutral Science? Analyzing the Legitimacy of Scientific Information in the Political Sphere." *Studies in History and Philosophy of Science Part A* 90 (December): 102–10. <https://doi.org/10.1016/j.shpsa.2021.08.009>.
- Lusk, Greg. 2025. "Science, Values, and Action Guidance: Can We Stop Talking about the Value Free Ideal?" In *Values, Pluralism, and Pragmatism: Themes from the Work of Matthew J. Brown*, edited by Jonathan Y. Tsou, Jamie Shaw, and Fehr Carla. Boston Studies in the Philosophy and History of Science. Springer.
- Lusk, Greg, and Kevin C. Elliott. 2022. "Non-Epistemic Values and Scientific Assessment: An Adequacy-for-Purpose View." *European Journal for Philosophy of Science* 12 (2): 35. <https://doi.org/10.1007/s13194-022-00458-w>.
- Lusk, Greg, and Kevin C. Elliott. 2024. "Against Global Aims for Science: Values, Epistemic Priority, and a Local Aims Approach." *Synthese* 204: 47. <https://doi.org/10.1007/s11229-024-04633-9>.
- McKie, Robin. 2019. "Climategate 10 Years on: What Lessons Have We Learned?" *The Observer*, November 9. <https://www.theguardian.com/theobserver/2019/nov/09/climategate-10-years-on-what-lessons-have-we-learned>.
- Mintz-Woo, Kian. 2022. "Carbon Pricing Ethics." *Philosophy Compass* 17 (1): e12803. <https://doi.org/10.1111/phc3.12803>.
- Nordhaus, William D. 2007. "A Review of the *Stern Review on the Economics of Climate Change*." *Journal of Economic Literature* 45 (3): 686–702. <https://doi.org/10.1257/jel.45.3.686>. <https://doi.org/10.1257/jel.45.3.686>.

- Oreskes, Naomi, and Erik M. Conway. 2010. *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming*. Bloomsbury.
- Parker, Wendy S. 2009. "Confirmation and Adequacy-for-Purpose in Climate Modelling." *Aristotelian Society Supplementary Volume* 83 (1): 233–49. <https://doi.org/10.1111/j.1467-8349.2009.00180.x>.
- Parker, Wendy S. 2010. "Scientific Models and Adequacy-for-Purpose." *The Modern Schoolman* 87 (3/4): 285–93. <https://doi.org/10.5840/schoolman2010873/410>.
- Parker, Wendy S. 2014. "Values and Uncertainties in Climate Prediction, Revisited." *Studies in History and Philosophy of Science Part A* 46 (June): 24–30. <https://doi.org/10.1016/j.shpsa.2013.11.003>.
- Parker, Wendy S. 2020. "Model Evaluation: An Adequacy-for-Purpose View." *Philosophy of Science* 87 (3): 457–77. <https://doi.org/10.1086/708691>.
- Parker, Wendy S., and Greg Lusk. 2019. "Incorporating User Values into Climate Services." *Bulletin of the American Meteorological Society* 100 (9): 1643–50. <https://doi.org/10.1175/BAMS-D-17-0325.1>.
- Rudner, Richard. 1953. "The Scientist qua Scientist Makes Value Judgments." *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Schroeder, Andrew S. 2026. "Scientific Knowledge as a Public Resource: Arguments and Challenges for a Democratic Approach to Values in Science." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Shabecoff, Philip. 1988. "Global Warming Has Begun, Expert Tells Senate." Section A. Page 1. *New York Times*, June 24.
- Stern, Nicholas. 2007. "Stern Review Final Report." UK. https://webarchive.nationalarchives.gov.uk/ukgwa/20100407172811/https://www.hm-treasury.gov.uk/stern_review_report.htm.
- Taper, Mark L., David F. Staples, and Bradley B. Shepard. 2008. "Model Structure Adequacy Analysis: Selecting Models on the Basis of Their Ability to Answer Scientific Questions." *Synthese* 163: 357–70. <https://doi.org/10.1007/s11229-007-9299-x>.
- Ward, Bob. 2023. "Newspapers Are Using Climate Researcher's False Claims about Journal Bias to Mislead Readers." Grantham Research Institute on Climate Change and the Environment. September 8. <https://www.lse.ac.uk/granthaminstitute/news/newspapers-are-using-climate-researchers-false-claims-about-journal-bias-to-mislead-readers/>.
- Ward, Zina B., and Kathleen A. Creel. 2024. "To Hedge or Not to Hedge: Scientific Claims and Public Justification." *Philosophy of Science*: 1–19. <https://doi.org/10.1017/psa.2024.17>.
- Weart, Spencer. 2025. "The Public and Climate Change Since 1980." The Discovery of Global Warming, April. <https://history.aip.org/climate/index.htm>.
- Winsberg, Eric. 2018. *Philosophy and Climate Science*. Cambridge University Press.

THE POWER OF VALUE-LADEN FRAMING

Examples from Wildlife Conservation Research

*Jacalyn M. Beck, Claire F. Hoffmann, Kelly Kapsar,
and Robert A. Montgomery*

43.1 Introduction

Value judgments are inherent in every scientific endeavor. As researchers, we express our values either consciously or unconsciously through the various frames we choose to use throughout the scientific process ([Elliott 2017](#)). Frames are communication techniques that define and draw attention to certain aspects of a complex system or scenario. Researchers use frames not only to communicate scientific results to others but also to conceptualize and structure investigations, determine arbitrary boundaries within study systems, declare problem statements, and identify and measure relevant variables. In the process of simplifying or isolating certain aspects of a system, researchers necessarily exclude others; thus, a frame is always a partial view of reality ([Woroniecki et al. 2020](#)). The choice of frames and framing techniques make a researcher's values detectable in ways that are measurable and may even lead to polarization and controversy within a field ([Vazquez et al. 2021](#)).

In this chapter, we will examine conservation biology, one scientific endeavor in which value judgments and value-laden framing are especially pertinent and have caused contention within and outside the research community. As four researchers with diverse experiences in international wildlife conservation, we have seen the expression and influence of values across all stages of the scientific process. Here, we analyze the significance of value-laden framing within our field, starting with a brief consideration of the historical and current narratives prevalent in conservation biology. We explain how such narratives ultimately hold power to influence values in relation to nature at both personal and global scales. Then, we outline major aspects of the research process and provide case studies from each of the authors' lived experiences navigating complicated value-laden scenarios. Finally, we conclude with suggestions on how researchers from any field can develop the skills needed to identify value judgments and address value-laden frames in their work.

43.2 The conservation context

Conservation biology is a mission-driven, and therefore value-based, discipline established around the goals of studying and protecting biodiversity. The field is entrenched in colonial history, whereby global research activities were often undertaken at the expense of Indigenous groups and traditional stewards of the land. The frames that researchers used in their work contributed to a broader conservation narrative espousing the value of nature (or particular aspects of it) over people (especially women and

rural, low-income, and BIPOC groups (Kareiva and Marvier 2012)). Characterized by the preservation of pristine landscapes, the creation of protected areas, and the reduction of species extinction risk, this narrative was dominant in conservation through the 1990s (Mace 2014). In the early 2000s, popular framing began to shift within the conservation literature and new ideas emerged around ecosystem services and the economic value of nature. Researchers began to frame nature as existing *for* people rather than *despite* people. This framing in turn led to a new prominent narrative that nature can and should be monetized and that people will be more likely to conserve nature if they receive financial benefits.

When prevailing narratives go unquestioned within a field, it becomes more difficult for researchers to publish or promote alternative or dissenting perspectives (Delborne 2016). Narratives can then take on their own strength to define reality, becoming indisputable discipline-specific discourses (Shackleton et al. 2023) with the power to influence behavior and advance certain actors and institutions (and their associated values) over others. For instance, research on the financial benefits of nature has sparked a broader movement encouraging payment for ecosystem services and monetary incentive programs to drive conservation agendas. This powerful capitalistic discourse has driven research and practice globally despite extensive evidence that oversimplified financial solutions do not resolve conservation conflicts (Lecuyer et al. 2019; Naughton-Treves, Grossberg, and Treves 2016). In this way, discipline-scale discourses, originating from collective researcher value judgments, can exercise and influence power both within the system under study and far beyond (Figure 43.1). That is, values are not only causal agents in the scientific process but are also indirect effects of it through the promotion and enforcement of certain norms (Ratti and Russo 2024).

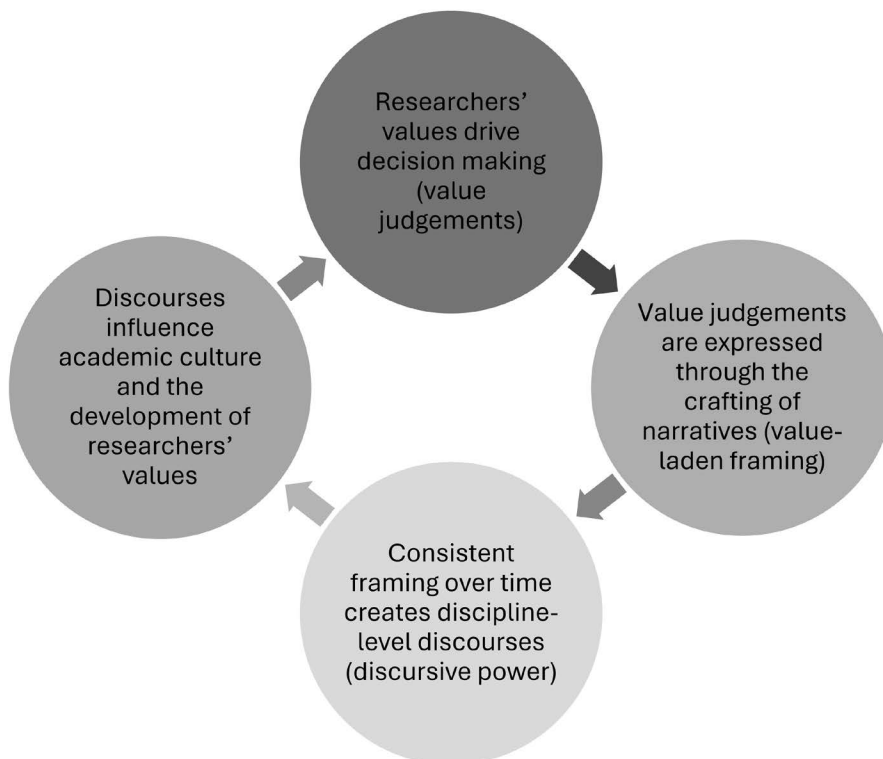


Figure 43.1 Researchers' Values both Affect and Are Affected by the Narratives and Discourses That Define Academic Culture.

In recent years, conservation research has been experiencing yet another paradigm shift in which explicit attention has been drawn to its exploitive, extractive history in hopes of rectifying past wrongs and incorporating more diverse perspectives and socially just research practices into the discipline (Montgomery et al. 2020). New frames are emerging that promote the interconnectedness of nature and people through the goals of building community resilience, leveraging socio-ecological systems theory, and practicing co-production and transdisciplinarity (Evans 2021). In a recent convening of nearly 300 multidisciplinary experts, a revised agenda for conservation research and practice was developed, with the first major recommendation being to revisit and refocus dominant conservation narratives (Wyborn et al. 2020). We propose here that researchers can meaningfully contribute to this aim by proactively attending to our own value judgments throughout the research process. We attempt to model this in the following sections and suggest that our case studies may serve as examples for scientists studying other complex systems and our lessons learned can be applied across disciplines.

43.3 Value-laden framing during research and analysis

Value judgments influence the topics we choose to study and the methods we use (Elliott 2017). For example, if statistical tests are used in data analysis, we implicitly communicate that mathematically significant results are more relevant to the system under study than other potential types of results. This is typical in conservation biology, which is traditionally grounded in positivist, reductionist methodologies. However, conservation researchers may be unaware that these choices are in fact value-laden because they represent tacit knowledge reinforced through years of teaching and training within the field (West and Schill 2022). Today, conservation is experiencing an increase in social science techniques and perspectives which can help researchers become more aware of the ethical-political dimensions of our chosen research methods (Bennett et al. 2017). However, the transition is not simple or linear, and the incorporation of these techniques by researchers takes practice and may be met with resistance (see Box 43.1).

Box 43.1 The unintended consequences of the status quo

Like many PhD students, I approached my first field season with a mixture of excitement and anxiety. I was to spend three months researching human-carnivore conflict in northern Tanzania, interviewing livestock owners to understand how carnivore behavior may influence their perceptions of risk. Working closely with a Tanzanian lab mate—fluent in Swahili and trained in social work—I felt completely prepared. I had read every possible relevant study, memorized site maps, and internalized the names of preeminent researchers. I thought I knew the system inside and out and had developed my interview questions based on that knowledge. However, this illusion started to crack during our very first interview. The *mzee* (elder) we spoke with was irritated and uninterested in our questions about livestock loss, instead fixating on reports about elephants eating his crops. I left uneasy, sensing I had missed something. Over the next few weeks, nearly every interview followed the same pattern. Then, one conversation shattered any lingering doubt I had.

At first, it was more of the same—the interviewee talked about elephants, not lions or hyenas. But soon he diverted into topics I couldn't follow with my mediocre Swahili. His wife abruptly interrupted, speaking harshly over him before storming off. An awkward silence followed. Finally, the *mzee* stood and led us toward the farm behind his *boma* (livestock enclosure). As we walked, my lab mate explained.

Their eldest son had been killed by an elephant in 2013 while defending their crops. He was only 22. The family reported the death to the authorities, who recorded the incident but never returned with compensation or assistance. Instead, a parade of foreign researchers arrived, asking about human-wildlife conflict and reopening their wound year after year. In one memorable moment, the mother said,

Why are you here? To make us relive his death again? To remind me of my dead son? We have told our story more times that I can count, and no one has ever helped us. Just leave us to mourn him in peace.

We hadn't intended to bring up their son's death; we hadn't even planned to discuss elephants. But I lacked awareness of the intersectionality of these issues. I had failed to see how livestock predation was deeply entwined with crop loss, and how those were inextricable from familial loss and distrust in both foreign researchers and the Tanzanian government. I had to see this broader picture through their trauma because it was not reflected in the literature in which I had immersed myself.

I realized that I had been unprepared to conduct that research effectively and ethically because the body of literature I relied on was framed within the values of those who had written, funded, and published it. More importantly, I had not yet learned to recognize how those framings shaped the narratives that defined my field. Western research institutions, publishers, and funding agencies are all biased toward a small suite of species, often charismatic and highly recognizable (Hoffmann and Montgomery 2022). As a result, our research system is largely compartmentalized into single-species studies, creating silos that obscure the interconnected nature of ecological and social dynamics, regardless of whether the researchers themselves want to remain siloed. This deep expertise on individual species can be valuable, but it has also led to a collective blind spot—an inability to fully grasp how human-wildlife conflicts compound and intersect (Lischka et al. 2018).

Since then, I have explored literature beyond these silos, uncovering work that more holistically examines socio-ecological interactions. I have read dozens of manuscripts calling for better integration of social and cultural considerations in human-wildlife conflict research. However, actually implementing these approaches can be slow and challenging, as we learned firsthand. After realizing we had inadvertently aligned with the overly narrow approaches represented in the literature, we were determined to correct that error in our analytical approach and framing of results. We applied a complex systems and grounded theory approach to explore both social and ecological aspects of human-wildlife conflict. However, during peer review, we encountered resistance. One reviewer dismissed our approach as “unnecessarily difficult and cumbersome.” Another questioned whether it was “applicable in reality” and whether it truly moved the needle in conservation science. Before we could verify our findings, we first had to justify the legitimacy of the approach itself. They asked why we couldn't just leave researchers to “divide and conquer” in their own areas of expertise, maintaining the comfortable status quo. My answer is simple: By perpetuating this pattern, we abdicate our responsibility for shaping the research culture that ultimately impacts the lives of the people and wildlife we interact with along the way.

This case illustrates how entrenched research frames—such as the narrow focus on carnivore conflict—can obscure broader socio-ecological realities, reinforcing incomplete narratives that shape both conservation practice and local perceptions of research.

43.4 Value-laden framing in science communication

The influence of value-laden framing may be most apparent in science communication and outreach. When writing and speaking about our work, researchers control the terms, definitions, and metaphors that go on to form the conservation narrative. Specific word choices reflect the authors' value orientations, biases, and motivations, and the use and popular meaning of words can change over time. For example, "natural resources," "wilderness," and "restoration" are all terms common in the conservation literature which have been questioned, scrutinized, and at times renounced, due to the values and histories the terms have come to convey (Corlett 2016; Ward 2019). Additionally, researchers may need to improve our knowledge of the historical contexts of Indigenous terminology, tribal affiliations, or proper place names to prevent the use of entrenched and harmful meanings and to acknowledge and avoid perpetuating injustices of colonial conservation (see Box 43.2).

Box 43.2 The power of a name

In the fall of 2017, I had just completed my first graduate research field season on interactions between African lions and Maasai cattle in northern Tanzania. Over several months, I walked nearly 500 miles following Maasai herders on their daily trek to fresh grass and water. I had been referring to the region as the "Maasai Steppe," as my academic collaborators had. But as I wrote up my research, I questioned whether this was the most appropriate term. None of my local assistants or guides used it. Seeking answers, I consulted colleagues and explored literature on Tanzanian and Maasai history, geography, and political ecology. Soon, the story emerged.

For centuries, East Africans have been murdered, enslaved, and evicted, notably by Portuguese colonists in the 1500s, Germans in the 1800s, and British in the 1900s. Colonists delineated stolen land for agriculture, mining, and hunting. In 1951, under British colonial rule, Serengeti was designated Tanzania's first national park. Soon after, game reserves were founded around Lake Manyara and the Tarangire River, securing fertile "wild" landscapes for Western tourists. This process necessitated forcibly removing pastoralist groups, including the Maasai, displacing them onto smaller, less arable parcels with restricted access to water. These constraints rendered their traditional semi-nomadic cattle grazing increasingly untenable. At the time, scientific discourse was dominated by the "tragedy of the commons," reinforcing colonial claims that communal land management led to overgrazing and environmental degradation (Nelson 2012). Conservationists and policymakers deemed pastoralism a threat to wildlife habitat, justifying further Maasai exclusion.

Yet, within decades, the narrative that framed Maasai practices as destructive began to shift. New ecological research challenged the overgrazing hypothesis, demonstrating that traditional pastoralist strategies were both sustainable and well-adapted to northern Tanzania's semi-arid landscape (Nelson 2012). But rather than returning Maasai land rights, conservationists—now concerned that Tarangire National Park was becoming too isolated by agricultural expansion—proposed a wildlife corridor between Tarangire and Lake Manyara (Bluwstein 2018). This required reclaiming even more land from local communities—many of whom had been forced into sedentary farming after their displacement. The same institutions that had once insisted the Maasai abandon their herding traditions now imposed strict regulations on livestock numbers and grazing practices under the guise of "community-based conservation." Conservation science kept shifting, but not in a way that restored Maasai autonomy. Tarangire itself came to represent an entire landscape threatened not by land theft or industry but by rural communities trying to adapt and survive.

In the 1990s, the international conservation NGO African Wildlife Foundation (AWF) further entrenched this dynamic. As part of a major fundraising push, AWF branded African environments as “AWF Heartlands,” positioning Tarangire as the core of the “Maasai-Steppe Heartland.” This name invoked the Maasai, not as an act of recognition but as a rhetorical device to attract donors. The branding campaign supported “threats-based conservation,” targeting the subsistence agriculture that colonial powers had previously forced upon the Maasai (Bluwstein 2018). In contrast, “acceptable” activities—those deemed compatible with conservation—were limited to tourism and research by Western institutions. AWF’s framing of the “conservation landscape” thus created spaces of inclusion for some and exclusion for others—expanding access only for conservationists, investors, and international visitors.

By the early 2000s, AWF’s influence extended beyond rhetoric—it took ownership of a strategic swath of land between Tarangire and Lake Manyara, establishing Manyara Ranch under the Tanzanian Land Conservation Trust. Initially presented as a community conservation area, it soon became clear to locals that this land no longer belonged to them. Ranch Management excluded the Maasai from decision-making, further stripping them of rights to determine how the land should be used (Goldman 2011). Early in my fieldwork, I planned to follow Maasai herders grazing in Manyara Ranch. However, I soon learned that they had been denied access to Ranch pastures that season. Although formal grazing requests were submitted, management determined the drought was not yet severe enough; only their Ranch-owned cattle were permitted to graze. As a result, herders were required to walk farther every day to find resources, often having to choose between good pastures and water. Sometimes cattle and herders alike went without water for days at a time.

In the months following my data collection, I pieced together the history of this land and my own position within it. Unlike my Maasai colleagues and friends, I could come and go as I pleased. They could not. Would it be appropriate to refer to this landscape using a name created by the very institutions that still controlled their lives and livelihoods? Ultimately, I used the term Tarangire-Manyara Ecosystem in my publications, framing the region by its natural features rather than the fleeting agendas of conservation elites. As of this writing, Maasai people are still labeled “invaders” and continue to be evicted from their traditional homelands in Tanzania (Weldemichel 2020). I contemplate now, as I did during graduate school: How much of Tanzania’s Maasailand will remain Maasai’s land?

The renaming of landscapes by conservation organizations exemplifies how framing is never neutral; it is an assertion of power that shapes who controls land, which voices are amplified, whose identities are legitimized—and whose are not.

In addition to specific words and phrases that researchers choose, the mode and style of communication, tone, voice, and imagery all influence the narrative while potentially promoting the researchers’ values. Researchers also implicitly make our values known by selecting target audiences for our work. While peer-reviewed articles are important in academic circles, it can be difficult to provide free access to this information or translate articles into media designed for private landowners, conservation practitioners, tribe members, policymakers, voting citizens, youth, or other non-academic audiences. Lead researchers have the power to determine who gets a say in the final product (e.g., graduate students, field technicians, research coordinators, local collaborators, interdisciplinary scholars, artists) and what that product ultimately looks or sounds like. Many researchers engage in outreach well beyond scientific publications and must consider how our value judgments will be received and whose worldviews we are endorsing (see Box 43.3).

Box 43.3 Joining voices for conservation outreach

Since 2015, I have been working with the Saint Louis Zoo on the co-development of interpretive messaging for the Zoo's McDonnell Polar Bear Point exhibit. When the exhibit was being conceived, my mentors at the Zoo wanted to be intentional about how climate change messaging to the public was framed. Many zoo exhibits tend to focus on species' ecology and natural history while overlooking the people who live with and often have close cultural ties to exhibited species. In the context of zoos' conservation-oriented missions, this portrayal of wildlife separate from humans can oversimplify complex issues and reinforce antiquated or impractical conservation approaches.

Specifically, polar bears are often portrayed as isolated species occupying an uninhabited Arctic. However, Indigenous peoples have lived alongside polar bears for thousands of years, identify with them as culturally important species, and sustainably harvest bears through subsistence hunting. While polar bears may be one of the most prevalent symbols of climate change in public media today, climate change poses threats not only to polar bears, but also to Indigenous peoples across the Arctic, and their intertwined ways of life. Communities have received harsh criticism for harvesting polar bears for both subsistence and cultural practices. Meanwhile, climate change caused primarily by non-Arctic communities and Western lifestyles is the primary threat to polar bears. Likewise, Indigenous peoples are experiencing dramatic negative impacts and facing rapid changes to their communities, livelihoods, and traditional ways of life ([Arctic Monitoring and Assessment Programme 2021](#)).

For our Zoo exhibit, we wanted to help visitors understand the complexities of human-wildlife coexistence in the Arctic, especially in the context of climate change. However, rather than writing text for the exhibit ourselves, my mentors and I collaborated with Indigenous communities of western Alaska to co-develop an interpretation method that would provide the communities with a metaphorical "megaphone" in the form of video journals. In their videos, Alaskan Native community members shared perspectives on their lives, their lives in relation to nature, and how both are changing in the modern world. In exchange for sharing this incredible firsthand perspective of Arctic life and climate change with the Zoo's audiences, the Zoo staff worked to fulfill the desires of the community to provide learning opportunities for their children. We collaboratively chose to hire a local teen media institute to teach students the skills needed to interview and film their elders for the creation of the video journals. This format resulted in not only the creation of videos shared with millions of visitors to the Zoo, but also the exchange of traditional knowledge between generations, and the education of students in journalism techniques. The final video journals are now owned by the respective tribes in which they were created and displayed with their permission at the Zoo's polar bear exhibit.

Through this novel approach to public outreach, we aimed to make progress toward rectifying several inaccuracies and injustices in the conservation community. First, we sought to upend traditional approaches to conservation which tend to oversimplify real-world complexities. The video journals serve to demonstrate the close relationships between polar bears and Indigenous communities and illustrate that the amplified rates of climate change affect not only the bears but also the humans that live alongside them. This portrayal is in contrast to oversimplified narratives of starving polar bears floating on ice bergs. Second, this approach served to connect Zoo audiences and Indigenous communities more directly by creating an opportunity for communities to determine their own framing and the overall message that they wanted to share with Zoo audiences, instead of more common approaches taken by zoos, such as sending outside film crews to craft a video of communities with which they are not familiar. Third, the project was co-developed between the Zoo and the Alaska Nanuuq Commission to be mutually beneficial, meeting

the interests and goals of both the Zoo (climate change outreach) and Indigenous communities (educational and traditional knowledge exchange opportunities for children). Finally, by using this approach we sought to ameliorate potential power imbalances by keeping the video journal ownership in the hands of the communities who are able to revoke the Zoo's ability to display them at any time. While this project is not perfect and will not solve systemic problems that have existed for centuries, we hope that it is a step in the right direction toward a more just and effective approach to conservation that celebrates and collaborates with the communities that have coexisted with wildlife for thousands of years.

By shifting from externally imposed frames to co-created ones, this project demonstrates how reflexivity and collaboration can help dismantle exclusionary narratives, creating space for multiple ways of knowing within conservation biology and public engagement.

43.5 Value-laden framing drives conservation policy and impact

The effects of value-laden framing go well beyond individual publications, public outreach, and broad disciplinary norms. Framing can have substantial impacts on international conservation strategies and conservation action on the ground. Researchers need to be aware of policymakers' values and select frames that successfully connect those values to our findings and the system under study (Pulkkinen et al. 2022). Funding sources also play a major role in determining conservation discourses. The value judgments of funding agencies shape the direction and motivations for scientific inquiry and influence how basic science principles are adopted and preserved over time within institutions (Salomon et al. 2023). It's critical for researchers to reflect on the power that funders have over not just our work but the future of conservation worldwide and how outcomes might change along with the framing of research priorities (see Box 43.4).

Box 43.4 Exploring fundamental versus applied science in a dynamic world

Over the course of my career, I have been fortunate to work at a diverse array of remote field sites around the world. One of my absolute favorites is in McMurdo Sound, Antarctica where the US Government has a permanent base. I spent just over three months living on the Ross Ice Shelf studying Weddell seal population dynamics. During that field season, I was confronted with a value dilemma involving how to frame the overall goal of my research efforts. I was pursuing population demography research questions in a landscape being rapidly reshaped by climate change. In exploring the roles of fundamental science (e.g., understanding basic principles) versus applied science (e.g., oriented to solve a particular problem), I delved into the context in which scientific research expanded in the world's most southerly continent.

By the middle of the twentieth century, the world's superpowers were freshly freed from the brutal conflicts defining WWII and could once again turn their attention to imperial conquests. With much of the world already conquered, subdivided, and ruled, Antarctica presented an appealingly blank map and military bases were established across Antarctica in the 1950s. In this case, the interests were obvious: The Antarctic likely held vast quantities of valuable precious metals, minerals, and hydrocarbons. The establishment of these bases rapidly escalated military tensions. The last thing that the world could endure was another war, and certainly not one waged in the most inhospitable place on Earth. Thus, in 1959 the US Government convened the countries participating in the International Geophysical Year for a discussion

of strategies and priorities over the Antarctic. Out of these meetings was borne the Antarctic Treaty which came into effect in 1961. The Treaty made clear that Antarctica would only be used for peaceful purposes, and while military presence was permitted, it would be only for scientific support and not for weapons testing or other military operations. Overnight McMurdo was rebranded from a military base run by the US Navy to a base for scientific study managed by the US National Science Foundation.

There are now over 75 bases on the Antarctic continent that deliver scientific research for scores of countries. The growth of the research (which is primarily fundamental in nature) has been staggering, with over 2,000 peer-reviewed papers on Antarctic science in just one year. Furthermore, supporting that science has been incredibly expensive. In 2023, the US Antarctic Program requested \$243 million of Congressional appropriations for Antarctic Facilities and Operations, \$77 million for Antarctic Logistical Support, and tens of millions more for individual Antarctic programs (“National Science Foundation FY 2023 Budget Request to Congress” 2022).

In spite of this research, the climate of Antarctica has warmed at a rate that is among the highest globally, the structure of the ice sheets and glaciers has destabilized, and populations of endemic species have rapidly declined (Eayrs et al. 2021). Thinking back to my study species, Weddell seals are among the most well-studied marine mammals in the world, and yet, the first circumpolar estimate of their populations was not completed until 2021 (LaRue et al. 2021). The delay in producing this type of estimate is, in part, attributable to a tendency for Antarctic science to be structured in fundamental rather than applied dimensions. Furthermore, the Weddell seal species population estimate returned from that research was some 30% lower than expectation. It is unclear whether previous predictions were inaccurate or whether the Weddell seal population is precipitously declining. Another complexity of this research is that it was based on satellite imagery that was already ten years old at the time of publication. Thus, if Weddell seal populations are declining, then prevailing science is presently far too reactive to document, much less respond to the challenge of restoring those populations.

Herein, the central value dilemma relates to the type of research that is supported. Fundamental research can undoubtedly unlock principles of the natural world with relevance to conservation and management. But it does not have obvious mechanisms to implement the outputs of that research process to positively restore ecosystems or species. Given the time, energy, and expense of administering the US Antarctic Program, I can’t help but wonder whether we would be better prepared to respond to the challenges of climate change and biodiversity loss if applied research had been similarly incentivized over the past 60+ years in the Antarctic. This is not a dilemma that is unique to the US Antarctic Program nor to the National Science Foundation. Given the severity of the coupled climate and biodiversity crisis for the sustainability of human society on Earth, there is a need for all funding bodies to scrutinize the research that they are supporting and to best align that research to solving these existential issues.

This case highlights how research priorities—driven by historical narratives, institutional norms, and funding structures—determine not only what knowledge is produced but also whether that knowledge is actionable in addressing urgent conservation challenges.

43.6 Advancing the narrative

In conservation biology, and other complex systems research, value-laden framing demands greater attention. Reflexivity is a critical first step in developing the self-awareness needed to identify and contend with hidden value judgments (Beck et al. 2021; Eyster, Satterfield, and Chan 2023). Through

reflexivity, researchers recognize our positionalities and epistemologies, exploring how they shift across contexts and experiences. In this way, reflexivity builds the foundation for relational thinking by helping researchers examine how our own perspectives intersect with others. In turn, relational thinking strengthens narrative-building by emphasizing emergent properties and dynamic system interactions (West et al. 2020). Together, these approaches prepare researchers to critically assess how their value judgments impact and influence others in the system, communicate across multiple frames, and apply different analytical lenses to evaluate their chosen narratives (Saltelli et al. 2020).

The evaluation of framing techniques, disciplinary narratives, and discursive power dynamics is best done collaboratively. Discussion groups, departmental seminars, philosophy book clubs, or dialog-based workshops can create space to deliberate on the ethical and political implications of value judgments and co-develop intentional, strategic framings that could move our field forward (West and Schill 2022). Research teams can also utilize available tools to craft Values Statements that formalize their commitments while increasing scientific credibility (Schoenle et al. 2024). Diversity in collaboration is essential to prevent dominant values from becoming entrenched and unchallenged within personal research circles or the discipline as a whole (Pulkkinen et al. 2022). Researchers should openly share lessons learned and mistakes made, critically evaluate “success stories,” and interrogate the design, interpretation, and presentation of our own and other’s work through multiple potential frames.

In this chapter, we’ve reflected on our own personal experiences navigating value-laden challenges, aiming to encourage greater reflexivity, transparency, and discussion around the role of framing within wildlife conservation and beyond. However, framing operates at multiple, interconnected levels—from the personal choices of individual researchers to the dominant narratives of entire disciplines and, ultimately, to the policy frameworks that shape conservation practice on the ground. While personal reflexivity and diverse collaborations can help us recognize and interrogate the value-laden assumptions embedded in our work, this work must be combined with broader structural efforts to become a true catalyst for change (Pienkowski et al. 2022). Science is shaped by institutional norms, funding priorities, and entrenched power dynamics that often reinforce existing narratives rather than challenge them. While individual researchers may adopt more intentional framing strategies, we must also push for systemic shifts—advocating for more inclusive publishing practices, interdisciplinary funding models, and policy engagement that allows for the deep consideration of diverse perspectives (Nyboer et al. 2021; Yoh et al. 2024). To truly advance our field, researchers must not only examine our own value-laden framing decisions but also work collectively to dismantle the rigid frames that have historically constrained conservation thought and practice.

References

- Arctic Monitoring and Assessment Programme. 2021. “Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-Makers.” <https://www.amap.no/documents/doc/arctic-climate-change-update-2021-key-trends-and-impacts.-summary-for-policy-makers/3508>.
- Beck, Jacalyn M., Kevin C. Elliott, Charlie R. Booher, Kristen A. Renn, and Robert A. Montgomery. 2021. “The Application of Reflexivity for Conservation Science.” *Biological Conservation* 62: 1–11. <https://doi.org/10.1016/j.biocon.2021.109322>.
- Bennett, Nathan J., Robin Roth, Sarah C. Klain, et al. 2017. “Conservation Social Science: Understanding and Integrating Human Dimensions to Improve Conservation.” *Biological Conservation* 205: 93–108. <https://doi.org/10.1016/j.biocon.2016.10.006>.
- Bluwstein, Jevgeniy. 2018. “From Colonial Fortresses to Neoliberal Landscapes in Northern Tanzania: A Biopolitical Ecology of Wildlife Conservation.” *Journal of Political Ecology* 25 (1): 144–68. <https://doi.org/10.2458/v25i1.22865>.
- Corlett, Richard T. 2016. “Restoration, Reintroduction, and Rewilding in a Changing World.” *Trends in Ecology and Evolution* 31 (6): 453–62. <https://doi.org/10.1016/j.tree.2016.02.017>.

- Delborne, Jason A. 2016. "Suppression and Dissent in Science." In *Handbook of Academic Integrity*, edited by Tracey Bretag. Springer. https://doi.org/10.1007/978-981-287-098-8_30.
- Eayrs, Clare, Xichen Li, Marilyn N. Raphael, and David M. Holland. 2021. "Rapid Decline in Antarctic Sea Ice in Recent Years Hints at Future Change." *Nature Geoscience* 14 (7): 460–64. <https://doi.org/10.1038/s41561-021-00768-3>.
- Elliott, Kevin C. 2017. *A Tapestry of Values: An Introduction to Values in Science*. Oxford University Press.
- Evans, Megan C. 2021. "Re-Conceptualizing the Role(s) of Science in Biodiversity Conservation." *Environmental Conservation* 48 (3): 151–60. <https://doi.org/10.1017/S0376892921000114>.
- Eyster, Harold N., Terre Satterfield, and Kai M. A. Chan. 2023. "Empirical Examples Demonstrate How Relational Thinking Might Enrich Science and Practice." *People and Nature* 5 (2): 455–69. <https://doi.org/10.1002/pan3.10453>.
- Goldman, Mara J. 2011. "Strangers in Their Own Land: Maasai and Wildlife Conservation in Northern Tanzania." *Conservation and Society* 9 (1): 65–79. <https://doi.org/10.4103/0972-4923.79194>.
- Hoffmann, Claire F., and Robert A. Montgomery. 2022. "Implications of Taxonomic Bias for Human-Carnivore Conflict Mitigation." *Oryx* 56 (6): 917–26. <https://doi.org/10.1017/S0030605321000582>.
- Kareiva, Peter, and Michelle Marvier. 2012. "What Is Conservation Science?" *BioScience* 62 (11): 962–69. <https://doi.org/10.1525/bio.2012.62.11.5>.
- LaRue, Michelle, Leo Salas, Nadav Nur, et al. 2021. "Insights from the First Global Population Estimate of Weddell Seals in Antarctica." *Science Advances* 7 (39). <https://doi.org/10.1126/sciadv.abh3674>.
- Lecuyer, Lou, Sophie Calm  , F. Guillaume Blanchet, Birgit Schmook, and Rehema M. White. 2019. "Factors Affecting Feelings of Justice in Biodiversity Conflicts: Toward Fairer Jaguar Management in Calakmul, Mexico." *Biological Conservation* 237 (July): 133–44. <https://doi.org/10.1016/j.biocon.2019.06.017>.
- Lischka, Stacy A., Tara L. Teel, Heather E. Johnson, et al. 2018. "A Conceptual Model for the Integration of Social and Ecological Information to Understand Human-Wildlife Interactions." *Biological Conservation* 225 (May): 80–87. <https://doi.org/10.1016/j.biocon.2018.06.020>.
- Mace, Georgina M. 2014. "Whose Conservation?" *Science* 345 (6204): 1558–60. <https://doi.org/10.1126/science.1254704>.
- Montgomery, Robert A., Kendi Borona, Herbert Kasozi, Tutilo Mudumba, and Mordecai Ogada. 2020. "Positioning Human Heritage at the Center of Conservation Practice." *Conservation Biology* 34 (5): 1122–30. <https://doi.org/10.1111/cobi.13483>.
- "National Science Foundation FY 2023 Budget Request to Congress." 2022. <https://www.nsf.gov/about/budget/fy2023/pdf/fy2023budget.pdf>.
- Naughton-Treves, Lisa, Rebecca Grossberg, and Adrian Treves. 2016. "Paying for Tolerance: Rural Citizens' Attitudes toward Wolf Depredation and Compensation." *Conservation Biology* 17 (6): 1500–11. <https://doi.org/10.1111/j.1523-1739.2003.00060.x>.
- Nelson, Fred. 2012. "Natural Conservationists? Evaluating the Impact of Pastoralist Land Use Practices on Tanzania's Wildlife Economy." *Pastoralism: Research, Policy and Practice* 2: 15. <https://doi.org/10.1186/2041-7136-2-15>.
- Nyboer, Elizabeth A., Vivian M. Nguyen, Nathan Young, et al. 2021. "Supporting Actionable Science for Environmental Policy: Advice for Funding Agencies from Decision Makers." *Frontiers in Conservation Science* 2: 693129. <https://doi.org/10.3389/fcsc.2021.693129>.
- Pienkowski, Thomas, Laur Kiik, Allison Catalano, et al. 2022. "Recognizing Reflexivity among Conservation Practitioners." *Conservation Biology* 37 (2): e14022. <https://doi.org/10.1111/cobi.14022>.
- Pulkkinen, Karoliina, Sabine Undorf, Frida Bender, et al. 2022. "The Value of Values in Climate Science." *Nature Climate Change* 12 (1): 4–6. <https://doi.org/10.1038/s41558-021-01238-9>.
- Ratti, Emanuele, and Federica Russo. 2024. "Science and Values: A Two-Way Direction." *European Journal for Philosophy of Science* 14 (1): 1–23. <https://doi.org/10.1007/s13194-024-00567-8>.
- Salomon, Anne K., Daniel K. Okamoto, Kii'iljuus Barbara J. Wilson, et al. 2023. "Disrupting and Diversifying the Values, Voices and Governance Principles That Shape Biodiversity Science and Management." *Philosophical Transactions of the Royal Society B: Biological Sciences* 378 (1881). <https://doi.org/10.1098/rstb.2022.0196>.
- Saltelli, Andrea, Lorenzo Benini, Silvio Funtowicz, et al. 2020. "The Technique is Never Neutral: How Methodological Choices Condition the Generation of Narratives for Sustainability." *Environmental Science & Policy* 106: 87–98. <https://doi.org/10.1016/j.envsci.2020.01.008>.
- Schoenle, Laura A., Caleb O'Brien, Jennifer Brousseau, and Amber Wendler. 2024. "Unleash the Power of Values: How to Conduct Better Science and Cultivate Thriving Research Groups." *Journal of Environmental Studies and Sciences* 15: 449–57. <https://doi.org/10.1007/s13412-024-00985-3>.
- Shackleton, Ross T., Gretchen Walters, Jevgeniy Bluwstein, et al. 2023. "Navigating Power in Conservation." *Conservation Science and Practice* 5 (3): 1–17. <https://doi.org/10.1111/csp2.12877>.

- Vazquez, Marco, Bronwyn McIlroy-Young, Daniel Steel, Amanda Giang, and Gunilla Öberg. 2021. "Exploring Scientists' Values by Analyzing How They Frame Nature and Uncertainty." *Risk Analysis* 41 (11): 2094–111. <https://doi.org/10.1111/risa.13701>.
- Ward, Kim. 2019. "For Wilderness or Wildness? Decolonising Rewilding." In *Rewilding*, edited by Nathalie Pettorelli, Sarah M. Durant, and Johan T. du Toit. Cambridge University Press. <https://doi.org/10.1017/9781108560962.003>.
- Weldemichel, Teklehaymanot G. 2020. "Othering Pastoralists, State Violence, and the Remaking of Boundaries in Tanzania's Militarised Wildlife Conservation Sector." *Antipode* 52 (5): 1496–518. <https://doi.org/10.1111/anti.12638>.
- West, Simon, and Caroline Schill. 2022. "Negotiating the Ethical-Political Dimensions of Research Methods: A Key Competency in Mixed Methods, Inter- and Transdisciplinary, and Co-Production Research." *Humanities and Social Sciences Communications* 9: 294. <https://doi.org/10.1057/s41599-022-01297-z>.
- West, Simon, L. Jamila Haider, Sanna Stålhammar, and Stephen Woroniecki. 2020. "A Relational Turn for Sustainability Science? Relational Thinking, Leverage Points and Transformations." *Ecosystems and People* 16 (1): 304–25. <https://doi.org/10.1080/26395916.2020.1814417>.
- Woroniecki, Stephen, Hausner Wendo, Ebba Brink, et al. 2020. "Nature Unsettled: How Knowledge and Power Shape 'Nature-Based' Approaches to Societal Challenges." *Global Environmental Change* 65: 102132. <https://doi.org/10.1016/j.gloenvcha.2020.102132>.
- Wyborn, C., J. Montana, N. Kalas, et al. 2020. "An Agenda for Research and Action toward Diverse and Just Futures for Life on Earth." *Conservation Biology* 35 (4): 1–12. <https://doi.org/10.1111/cobi.13671>.
- Yoh, Natalie, Mukhlish Jamal Musa Holle, Jasmin Willis, Lauren F. Rudd, Iain M. Fraser, and Diogo Veríssimo. 2024. "Understanding Author Choices in the Current Conservation Publishing Landscape." *Conservation Biology* 39 (2): 1–10. <https://doi.org/10.1111/cobi.14369>.

VALUES IN ECONOMICS

Elizabeth Anderson

44.1 Contrasting values in neoclassical and classical economics

The dominant values of neoclassical economics today are economic growth and efficiency. Economists measure economic growth in terms of growth rates of total or average GDP. They often use average GDP as a measure of the average welfare of a country's population. Economists also prioritize two types of efficiency: productive and allocative. An economic state is productively efficient if each good and service is produced at the lowest possible average cost. An economic state is allocatively (Pareto) efficient if all opportunities to improve the well-being of one person without decreasing the well-being of anyone else have been taken. According to the first theorem of welfare economics ([Arrow 1951](#)), Pareto efficiency is achieved when (assuming no externalities, asymmetrical information, or monopolies) the markets for every good and service are in equilibrium. That economic growth and efficiency are valuable rests on the assumption that individuals improve their well-being in attaining a state they prefer more. When economists operationalize average well-being in terms of the growth of average GDP, they assume (among other things) that people prefer more income over less.

Economists have claimed that their preference-based definition of well-being is value-neutral because it avoids controversial claims: it refuses to judge the content of different preferences, or to make interpersonal comparisons of well-being ([Robbins 1948](#), 24–26, 144–52). Yet, [as Alexandrova \(2018\)](#) has argued, claims about what promotes well-being are “mixed”: they combine empirical claims with prudential, moral, or political value judgments. To avoid imposing values on people who have good reason to reject them, economists need to explore the implications of their definitions and measures of well-being and open them to public contestation.

We do have ethical reasons to question these implications. In focusing on average outcomes, economists disregard considerations of distributive justice and imply that increasing inequality does not harm anyone. In prioritizing economic growth, they assume that we are better off with higher consumption. In assuming that people typically advance their welfare through consumption rather than work, they accept as given the intrinsic disutility and merely instrumental value of work to workers and tend to think that making the experience or conditions of work better is less efficient than compensating workers with higher pay ([Anderson 2017](#), 111). In prioritizing productive efficiency, economists also tend to regard firms that manage to lower wages as doing something good, since this lowers the cost of production. By contrast, they don't generally regard higher profits as increasing the cost of production but rather as a sign that firms are delivering unexpectedly high value to consumers.

Economists today could learn from their classical predecessors, who made the ethical and political implications of their definitions of well-being explicit, opened them to public contestation, and used their values to guide their measures of welfare and choices of questions to investigate. Leading classical economists argued about distributive justice. Some thought that the most important consequences of economic institutions were not consumer goods and services, but our capacities, dispositions, and social relationships. They wanted work to involve activities that develop our agency, talents, and virtues. They wanted economic institutions to put people into relations of cooperation and reciprocal sympathy, as opposed to antagonistic competition and envy. They wanted these arrangements to dignify workers and enable them to think and stand up for themselves.

All the values I have mentioned are rooted in the Protestant work ethic. Neoclassical economists adopted only some of them. By studying the ethical debates of classical economists, we can recover a richer range of values relevant for twenty-first-century economics.

44.2 The values of the Protestant work ethic¹

The Protestant work ethic was invented by Puritan (Calvinist) theologians in sixteenth- and seventeenth-century England. Calvinists hold that salvation is by faith alone, which is an undeserved gift from God. They also affirm predestination, which implies that there is nothing we can do to change our fate in the next life. These doctrines made people desperate to learn whether they were saved, while removing damnation as a deterrent to sin. Calvinists solved these problems by arguing that a person's faith is revealed solely through their actions in obedience to God's will. God prescribes a utilitarian ethic to humans: he commands each of us to do our part to maximize the well-being of humanity. "Our part" is defined by our occupation in the division of labor. Because Calvinist theology rejects the possibility of supererogation, everything we do must be a performance of duty. Hence, to be sure of our salvation, we must follow God's will all the time.

The fundamental virtues of the work ethic follow from these doctrines. (1) Industry: we must work as hard as we can. (2) Frugality: we must waste nothing. We must put everything—natural resources, time, and financial assets—to good use. (3) Asceticism: we should not indulge in luxury, licentiousness, drinking, or other useless or harmful pleasures. (4) Contentment: we should not envy what other people have. (5) We should make as much money as we can through honest work that helps other people. Since wealth must not go to waste, after spending modestly for the comfort of oneself and one's own family, we must dedicate the rest to helping other people, either currently through charity or public works, or in the future through investment in honest businesses.

This work ethic was class-neutral. Every able-bodied person, whether rich or poor, must work. God calls everyone to a particular occupation that suits them. Because individuals are the best judges of what suits them, everyone has a right to free choice of occupation. Everyone relies on workers performing other roles in the division of labor, which God has instituted for the benefit of everyone. All honest occupations are therefore equal, and every diligent worker is entitled to respect. Every worker is also entitled to living wages, safe working conditions, and freedom from tyrannical abuse by their superiors. God willed that social arrangements protect and promote the life of every particular person in society. Hence, those unable to work have a right to charity. Puritans condemned exploitative business models, including monopolies, usury, and price discrimination that takes advantage of poor or desperate people. They also condemned ways of making money, such as gambling and pure speculation, that don't add value.

44.3 How work ethic values influenced classical economics

As the Industrial Revolution spread during the late eighteenth and nineteenth centuries, the work ethic secularized. Prudence replaced assurance of salvation as the prime motive for following the work ethic (Weber 2003, 11–4). Constraints on luxury and leisure were relaxed. The Industrial Revolution drew far sharper lines between classes. The model workers for seventeenth-century Puritans were craftsmen and yeoman farmers: manual workers who also enjoyed capital returns through their possession of tools, shops, or land. During the Industrial Revolution, both classes lost their capital and became impoverished wage laborers in large farms and factories. The work ethic divided into two versions. What I call the “conservative” work ethic reflected the interests and perspectives of capitalists and landowners. The “progressive” work ethic reflected the interests and perspectives of wage workers. Both versions of the work ethic informed the debates among classical economists over property rights, work regimes, and social welfare policies (Anderson 2023, chap. 3).

Classical economists agreed on three ideas. First, economics aims to determine which institutions and policies would best promote human welfare. Second, work ethic virtues are critical to the promotion of human welfare. Third, individuals change their behavior in response to incentives. Many of their core disagreements turned on the following question: which institutional arrangements would best promote work ethic virtues, by giving people incentives to practice them? This was largely an empirical dispute.

However, behind their disagreement lay two ethical questions. First, does human flourishing depend primarily on our consumption, or also on the quality of our social relationships and the exercise of autonomy and talents in all domains of life, including work? Second, should institutions be arranged to ensure the flourishing of every particular member of society? Or could some people be sacrificed for the sake of promoting greater aggregate welfare? That is, should we regard the promotion of human welfare as including a distributive requirement? If people care about their autonomy, exercising talents, the quality of their social relationships, and whether they are being treated well by others, their prospects for attaining these goods will affect their dispositions to practice the work ethic. So the empirical question is not entirely separated from the ethical question, because ethical ideas shape our conceptions of the good and hence our responses to different incentives.

The core empirical disagreement between economists who presupposed a conservative or progressive work ethic was whether individuals responded to incentives differently depending on their class status. Progressives claimed that everyone responded alike to incentives. Adam Smith (1776 [1981]a I.11.d.1; II.3.36; III.2.8–13) argued that people work harder to the extent that doing so will improve their condition. So, if capitalists want workers to work harder, they should pay them higher wages. The self-employed work even harder than wage laborers, because they get to keep all the fruits of their labors.

Conservatives thought that the poor—the majority of workers—were lazy, licentious spendthrifts. They wasted their money on drink, had more children than they could support, and would quit work as soon as they earned their subsistence. Raising their wage rate would only induce them to quit work earlier. They would not produce the surplus required to support the upper classes, who were responsible for the arts, sciences, religion, and everything else that makes society civilized. To induce them to work diligently for long hours, avoid wasteful spending, and restrain their reproduction, wages must be kept low and employment uncertain. This argument justified wage labor over self-employment. Wage workers work harder than the self-employed because their income is lower and they must work at a pace set by their boss. Conservatives preached asceticism and contentment to wage workers, as the original work ethic had demanded of everyone (Anderson 2023, 88, 95–6, 116, 127).

Conservatives held that the middle classes—managers, professionals, business owners—responded differently from the poor to economic opportunity. Envy of each others’ conspicuous

consumption of luxuries would spur the middle classes to work harder to win the race for superior wealth (Anderson 2023, 72, 96–7).

Conservatives thought the work ethic barely applied to aristocrats and landed gentry. A leisure class was needed to uphold civilization. Some would engage in the arts and sciences. Others would join the clergy, obtain a military commission, or serve in a reputable government office. None of these activities were seen as particularly demanding of effort, time, or attention (Anderson 2023, 87).

The conservative work ethic rationalized the three-class order of capitalist England. It purported to maximize human well-being, even as it justified subjecting the majority to poverty and oppressive work conditions. Conservatives justified their position in two ways. Malthus argued that the imprudent poor had to be sacrificed to promote aggregate well-being. Poverty relief incentivized vice. If they expected relief, the poor would have more children than they could support, thereby increasing the population, reducing wage rates, and creating even more poor people. Abolish relief and the poor would be forced to suffer the natural consequences of their imprudence. This would either teach them to internalize the work ethic or kill off their children. Either way, the survivors would follow the work ethic and thereby supply the labor needed to promote aggregate well-being (Malthus 1798).

Malthus's justification for sacrificing the poor altered utilitarianism by expunging distributive concerns from its foundations. His perspective led to the version of utilitarianism familiar today, which focuses on maximizing total or average utility, permits sacrificing some for the sake of greater total welfare, and allows distributive concerns to enter only contingently.

Bentham, although he also held a purely aggregative utilitarian view, rejected Malthus's claim that the poor must be sacrificed to promote total happiness. Rather than leaving the improvident poor to starve, they should be imprisoned in workhouses and forced to learn the work ethic through totalitarian discipline. In his workhouse plan, inmates would enjoy no civil or personal liberties. They would be isolated in prison cells and consigned to drudgery for the maximum profit of their keepers. Husbands would be separated from wives, men from women, and children from parents. Yet inmates could be happy even when stripped of dignity, privacy, freedom, and social relationships, because his ideal workhouse would offer the poor all they needed for happiness: food, shelter, clothing, and health care. His reduction of happiness to material consumption followed from his view that pleasures differ only in quantity and not quality and hence are fully substitutable for each other (Bentham 1812).

John Stuart Mill, a leading classical economist of the progressive work ethic, rejected Bentham's view that pleasure is homogeneous. There are higher and lower pleasures. People acquainted with both types of pleasure will not willingly give up the higher pleasures for any quantity of the lower. Hence, these pleasures are not fully substitutable. Higher pleasures are the pleasures obtained from the activities of higher or distinctively human faculties. Chief among these higher faculties are autonomy and sympathy (Mill 1985, 210–3, 231–3).

Mill's conception of human welfare led him to reject Malthus's and Bentham's proposals. If the poor lack work ethic virtues, this is because they have neither acquaintance with nor opportunities for higher pleasures. Wage labor provided no such opportunities, as it consisted in mindless and exhausting drudgery for bare subsistence. To end poverty, the poor must be able to experience and enhance their opportunities for higher pleasures. If an entire generation of poor children were provided a free education and a decent standard of living, they would learn how to restrain their impulses, exercise prudence, and acquaint themselves with pleasures beyond subsistence. This would spur them to internalize the work ethic (Mill 1965a, 372–5). Their willingness to supply labor would then increase with the wage rate.

Mill followed Smith in holding that the self-employed work harder than wage laborers. Yeomen are more productive than farmhands. Both advocated policies to promote peasant proprietorship. Due to economies of scale, factories could not be efficiently divided into small workshops. Mill (1965b, Bk. IV, chap. 7) argued that factory workers, out of desires for autonomy, would jointly build or purchase factories and operate them as workers' cooperatives.

Mill deplored the conservative model of progress based on stoking envy and competitive consumption because it promoted antagonistic social relations. He welcomed the arrival of David Ricardo's theorized "stationary state," at which economic growth would cease, because such a state would eliminate incentives for people to scramble over each other in the race for luxury. Instead, workers would have more leisure, where they would enjoy the higher pleasures of sympathetic association with family and friends. Sympathy is cultivated in social relations of equality, including workers' co-operatives. Mill identified sympathy, or the "feeling of unity with others," as the central moral motive for following a utilitarian ethic. Sympathetic people incorporate others' welfare into their conception of their own welfare. Mill thus continued the commitment of the original work ethic to promote the welfare of every particular person. No one should be sacrificed for the sake of a purportedly greater aggregate welfare ([Anderson 2023](#), chap. 6).

Other economists of the progressive work ethic also incorporated distributive concerns into their economic arguments. Adam Smith thought that it was far more important to raise wages than profits, since workers were by far the most numerous class. David [Ricardo \(1951\)](#) thought that the central question of economics was understanding the distribution of income by class (6). Public policy should focus on improving the condition and prospects of the working class, who are "by far the most important class of society" ([Ricardo 1951](#), 425). This was the original idea behind the utilitarian demand to promote "the greatest happiness of the greatest number." The workers are "the greatest number." The proper aim of public policy is their welfare, not that of the rich.²

Classical economists of the progressive work ethic from [Smith \(1776 \[1981\]b, V.f.50–61\)](#) and [Mill \(1965b, Bk. IV, chap. 7\)](#) to [Marx \(1975\)](#) deplored the ways factory work undermined workers' capabilities and social relationships. The micro-division of labor in factories consigned workers to tedious repetitive motions under the harsh discipline of their bosses. Workers were thereby deprived of skills, autonomy, dignity, and interesting work. Smith, Mill, and Marx agreed that factory labor demeaned workers, lowered their sights, undermined their health, and fostered class conflict. Work should be reorganized to enhance the autonomy and capabilities of workers and promote positive social relations of cooperation, sympathy, and solidarity. If workers controlled the organization of work, they would neither choose the micro-division of labor nor its dangerous conditions. Smith, Mill, and Marx favored either self-employment and/or workers' collectives to enable workers to choose how to organize work and control its conditions.

44.4 Neoclassical and heterodox economic perspectives on values of and in work

By the mid-twentieth century, neoclassical economists assumed that economic actors had internalized the work ethic. Business owners make as much money (profits) as they can. This demands frugality, or maximizing productive efficiency. The assumption that maximizing productive efficiency is good omits from the welfare calculus any direct consideration of how workers experience work, of the impact of work on workers' capabilities, or on the character of social relations at work. Neoclassical economics thereby adopts the perspective of the conservative work ethic. It implicitly accepts the Industrial Revolution's sharp division of the domains of work and leisure/consumption. In the workplace, everything including workers is viewed as purely an instrument—a "factor of production"—deployed as required to maximize productivity. Whether workers are exhausted, bored, deskilled, or demeaned at work is irrelevant. Only in leisure hours do workers get to be treated as ends—as consumers whose interests must be catered to.

The neoclassical framework allows a minor exception to this assumption in the theory of compensating differentials. Workers who want safer, more pleasant work conditions or more dignity and autonomy at work may be able to trade off some pay to obtain these goods. They take part of their

compensation in the form of in-kind goods enjoyed at work rather than income spent on consumption after hours. Either way, equally skilled workers competing in the same labor market will enjoy equal total compensation. All goods are substitutable at the right price. Neoclassical economists use the theory of compensating differentials to justify bad working conditions by arguing that workers regard themselves as fully compensated by their higher pay for enduring them (Viscusi 1983).

Heterodox economists who follow the progressive work ethic tradition dispute this analysis. This tradition includes many normative concerns of the original work ethic omitted by the conservative work ethic, including the rights to safe working conditions and living wages, and opposition to workplace tyranny and business models that take advantage of others' vulnerability or sacrifice the vital interests of some for the pleasure and convenience of others.

Larry Mishel and colleagues (2022) reject the theory of compensating differentials because workers have unequal bargaining power with their employers. Their weak bargaining position makes them unable to negotiate compensating differentials. This is why the worst-paid jobs also tend to be the most degrading, deskilled, monotonous, exhausting, and unhealthy jobs. Workers are forced to accept these conditions for lack of reasonable alternatives.

Marxist Harry Braverman (1998) challenges the deskilled micro-division of labor and authoritarian social relations between managers and workers in capitalist firms. He argues that managers design the firm's division of labor to deskill most jobs. This maximizes workers' replaceability and thereby minimizes their bargaining power. Yet when workers collectively own their firm, they use their power to divide the labor in ways that protect everyone's autonomy and skills, and without subjecting themselves to tedious drudgery. Moreover, these workers' cooperatives are at least equally productive as capitalist firms.

Capitalists' decisions to micro-divide labor are therefore not needed for efficiency but adopted to maximize profits by reducing labor's share of the firm's revenues. Either way of dividing the labor might be equally productively efficient when costs are judged narrowly, excluding the costs inflicted on workers by the micro-division of labor. Yet from a broader human welfare perspective, profit-maximizing firms promote human welfare less effectively than firms that organize the division of labor as workers' cooperatives do. This possibility is obscured by the default representation of wages but not profits as a cost of production. In addition, the assumption that there is a necessary tradeoff between work quality and pay obscures the possibility that profits may be obtained by worsening workers' bargaining power instead of increasing efficiency.

In the service sector, the assumption that the reduction of workers to interchangeable parts reflects the demands of maximizing productivity presupposes additional questionable value judgments about consumer welfare. Service workers are often turned into interchangeable parts by simplifying and routinizing their work through protocols and scripts, which remove judgment from the job. This enables a low-wage, high-paced job design with high turnover due to exhaustion and misery. Such modes of service provision assume that enduring relationships between particular service providers and particular recipients—e.g., between teachers and students, doctors and patients, social workers and clients—make no difference to human welfare. They also assume that customers' needs are homogeneous and can be addressed by workers deprived of autonomy. In practice, this means that time spent addressing idiosyncratic emotional needs, such as comforting the sick, helping a student manage conflict with peers, or building connections with other community members to help clients overcome loneliness, is time wasted. These values of personal interaction, which cannot be standardized but require attunement to the particular individual being served, are the first to be sacrificed by profit-maximizing firms when they increase quotas for how many people they serve in a given time period. These assumptions about the welfare of workers and customers, which underlie seemingly more objective claims about efficiency, illustrate Longino's (1990) argument that value-laden background assumptions affect how scientists judge whether the evidence supports their hypotheses.

The neoclassical response to this argument replicates the logic of compensating wage differentials. If customers want better than routinized, deskilled service, they will be willing to pay for it. Indeed, the rich do pay for tutors for their children, concierge medicine, and similar catered services. However, to the extent that turning workers into interchangeable parts redistributes income from labor to capital, economic inequality will increase, because inequalities in capital ownership greatly exceed inequalities in labor income. As more service workers are deskilled, their wages drop.

Such low-wage workers will not be able to afford more responsive and skilled services. *The economic system will be more and more oriented to the greatest happiness of fewer people, at the expense of the greatest number.* This is why progressive classical economists refused to reduce “the greatest happiness of the greatest number” to “the greatest total happiness.” That reduction was a conservative strategy to remove the distributive requirement inherent in the original Puritan work ethic, to ensure that “every particular person” be cared for.

44.5 The progressive work ethic and the capabilities approach to human welfare

We can partially describe the dispute between the neoclassical and classical progressive perspectives in terms of a disagreement over appropriate welfare metrics.³ Mid-twentieth-century neoclassical economists claimed that the metric of market prices, which reflect individuals’ preferences in terms of their willingness to pay for various goods, was neutral about the value of different individuals’ preferences and hence avoided irresolvable conflicts over value judgments (Robbins 1948, 24–6, 144–52). By contrast, if welfare metrics include such items as opportunities to exercise skills at work and positive social relationships, singling out such values for special consideration would seem to impose value judgments on individuals who prefer other goods.

Against these claims, I join the heterodox economists in arguing that the unequal bargaining power of workers relative to employers, and of ordinary consumers relative to the rich as they compete for services, reflects a grossly unjust weighting of the interests of the wealthy relative to everyone else. Indeed, on standard neoclassical assumptions, efficient markets give to “each individual a weight inversely proportional to his or her marginal utility of wealth” (DeLong 2009). Ordinary workers’ and consumers’ concerns are reduced to practical insignificance relative to the top 1% on such a weighting. Moreover, reliance on markets driven by conservative work ethic values generates negative externalities for social relationships. This includes a society riven by envy.

Suppose we evaluated societies instead from the perspective of the quality of social relationships and of the abilities of the people living in them, rather than just in terms of their consumption possibilities. The capability approach to measuring welfare (Sen 1999) offers a useful way to incorporate these broader concerns into our assessments. Call “functionings” the states of being and doing that are good for people, such as being adequately nourished and sheltered, and being able to move about freely in society. Capabilities are freedoms (real opportunities) to attain or exercise functionings, accessible by deploying the physical, mental, material, and social resources at one’s disposal. By focusing on freedoms rather than attainments, the capability approach avoids the objection that not everyone wants the functioning in question. Because the functionings at issue are attainable, they avoid the objection to average income metrics that they fail to address the concerns of those for whom the average is unattainable.

The capability approach has many applications. The United Nations uses it to construct the Human Development Index (HDI) to measure a country’s level of economic development. In addition to the standard measure of per capital income, the HDI also includes measures of education and health attainment, which do not perfectly track average income but are independently valuable.

The capability approach permits us to choose which dimensions of human welfare to focus on, and what questions to ask. We can use it to investigate the externalities of institutions that create more or

less social and economic inequality. Amartya Sen (1999, 73–4, 89), for example, has argued, following Smith, that the ability to appear in public without shame is relative to what others are typically able to afford: the cost of a respectable wardrobe increases with the standard of living. Hence a relative deprivation of income can create an absolute deprivation of capability. Nussbaum (2000) lists many functionings critical to human flourishing that reflect the quality of our social relationships, and which are harder for women to attain in more sexist societies. These include being recognized as a dignified person, freedom to form friendships and other affiliations, and security against violence, discrimination, stigmatization, and humiliation.

Thomas Piketty (2014, 298–302) explores the relational externalities of socioeconomic inequality by reviving Ricardo's focus on a class-based analysis of the distribution of income and wealth. He traces the social consequences of the growth in the capital share of national income as the ratio of total wealth to national income increases. As inequality between capital owners and workers increases to extremes, people behave less helpfully and become more manipulative and even criminal. By the early nineteenth century, the capital share of income was so great, and wealth so concentrated in a few hands, that honest working was not a viable path of class mobility. As Austen and Balzac illustrated in their novels, to get to the top of society, one would either have to be born into a wealthy family, marry into one, or acquire great wealth illicitly. Piketty warns that if current trends of concentrated wealth accumulation continue, today's rich societies are likely to face the same problem. A few families of vast dynastic wealth will dominate everyone else, spread a culture of disdain for those who must work honestly for a living, and divert ambitious people away from productive work to marital scheming or criminal activity.

Let's return to the concerns of Section 4, regarding workers' access to skilled, interesting, respected, and well-paid jobs, and how such jobs affect the quality of services to consumers. The development of every child's capabilities depends on adults caring for them in sustained relationships. Contra Bentham, child development requires much more than being raised in an orphanage staffed by interchangeable workers who provide only material subsistence. This fact has at least two implications for economics. One concerns national income accounting. Women perform most childcare work. Yet, if they do so as unpaid family members, their work is not counted in the GDP. Feminist economists have rightly objected to this exclusion as arbitrary, sexist, and a reflection of the ways capitalism exploits women's unpaid labor of social reproduction. The Puritan inventors of the work ethic, by contrast, classified women engaged in child-rearing as educators and recognized their work as socially necessary even when it was unpaid (Anderson 2023, 12). Under current accounting rules, paid educational labor is counted, but only as consumption, not investment. Yet even in narrowly economic terms, caring for children develops their human capital—the abilities needed to engage in future productive work. Hence, this labor should count as investment. Transfer payments needed to sustain both children and the adults who care for them should therefore also count as investment rather than excluded from GDP, as current accounting rules require (Block 2025, chap. 5).

Second, the centrality of sustained caring relationships to capability development and maintenance has important implications for the design of social services work. Economist William Beveridge (1942) planned the modern British welfare state on a factory model: interchangeable social services workers would deliver standardized goods and services to the population. In the 1990s, much of this work was outsourced to for-profit firms that focus on minimizing costs by routinizing the work while increasing service quotas per worker. This has predictably led to declining service quality, as well as numerous welfare recipients mired in complex problems including chronic unemployment, school failure, and juvenile delinquency. Social service workers have experienced burnout due to their inability to ameliorate such problems despite massive intervention.

Hilary Cottam (2019), a social entrepreneur and innovator in social services design, argues that the Beveridge model fails to address today's social problems because it is built on a transactional model

of service delivery addressed to managing need, with a focus on efficiency and productivity measures based on following protocols rather than creative problem-solving. Her radical experiments in remodeling social services focus rather on developing recipients' capabilities to manage their own lives by enlisting service workers from multiple agencies to team up, build relationships with recipients, and help them envision and pursue a better future. She documents how her model succeeds in achieving more ambitious goals than the Beveridge model. It does so at lower long-term cost, because investing in recipients' capabilities reduces their future reliance on welfare services. Moreover, the workers in these redesigned jobs exercise higher skills and autonomy and find their work more fulfilling.

44.6 Conclusion

I have argued that neoclassical economists' narrow focus on efficiency and economic growth incorporates problematic assumptions about human welfare and its measurement into their assessments of economic arrangements. They neglect core values of progressive classical economists such as Smith, Mill, and Marx because they (sometimes unwittingly) followed the path of conservative economists who wanted to maintain a class hierarchy in which property owners kept workers in poverty, precarity, and subordination. We should reject this implicit class bias and pay attention to the negative effects on workers of work designs that turn workers into interchangeable parts with little bargaining power by consigning workers to low-paid, deskilled drudgery. The capabilities approach to measuring welfare allows us to expand the normative vision of economics to include key concerns of progressive economists in enhancing human capabilities and promoting social relationships of equality and sympathy rather than envy and exploitation. Such an approach enables us to envision a redesign of service work that enhances the capabilities of workers and consumers alike, without sacrificing productive efficiency. My argument illustrates the power of Alexandrova's (2018) framework for assessing claims about objectivity in the science of well-being.

Notes

- 1 My discussion in Section 44.2 follows Anderson (2023, chap. 1). Weber (2003) offers the classic account, which I argue captures only the conservative version of the work ethic.
- 2 Thanks to Don Herzog for this insight.
- 3 See the contribution on values in measurement by C. Larroulet Philippi in this volume (Larroulet Philippi 2026).

References

- Alexandrova, Anna. 2018. "Can the Science of Well-Being Be Objective?" *British Journal for the Philosophy of Science* 69 (2): 421–45. <https://doi.org/10.1093/bjps/axw027>.
- Anderson, Elizabeth. 2017. *Private Government: How Employers Rule our Lives (And Why We Don't Talk About It)*. Edited by Stephen Macedo. Princeton University Press.
- Anderson, Elizabeth. 2023. *Hijacked: How Neoliberalism Turned the Work Ethic Against Workers and How Workers Can Take It Back*. Cambridge University Press.
- Arrow, Kenneth. 1951. "An Extension of the Basic Theorems of Classical Welfare Economics." In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, Vol. 2*, edited by Jerzy Neyman. University of California Press.
- Bentham, Jeremy. 1812. *Pauper Management Improved: Particularly by Means of an Application of the Panopticon Principle of Construction*. R. Baldwin.
- Beveridge, William. 1942. *Social Insurance and Allied Services*. HMSO.
- Block, Fred. 2025. *The Habitation Society: Creating Sustainable Prosperity*. Agenda Publishing.
- Braverman, Harry. 1998. *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. Monthly Review Press.

- Cottam, Hilary. 2019. *Radical Help: How We Can Remake the Relationships Between Us and Revolutionize the Welfare State*. Virago.
- DeLong, Bradford. 2009. "Hoisted from the Archives: A Non-Socratic Dialogue on Social Welfare Functions." *Brad DeLong's Semi-Daily Journal*, April 13. <http://delong.typepad.com/sdj/2009/04/hoisted-from-the-archives-a-non-socratic-dialogue-on-social-welfare-functions.html>.
- Larroulet Philippi, Cristian. 2026. "Values and Measurement." In *The Routledge Handbook on Values in Science*, edited by Kevin Elliott and Ted Richards. Routledge.
- Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton University Press. <https://doi.org/10.1515/9780691209753>.
- Malthus, Thomas. 1798. *An Essay on the Principle of Population*. Printed for J. Johnson. <https://oll.libertyfund.org/titles/311>.
- Marx, Karl. 1975. "Economic and Philosophic Manuscripts of 1844." In *Marx Engels Collected Works, Vol. 3*. Lawrence & Wishart.
- Mill, John Stuart. 1965a. *Collected Works of John Stuart Mill: Vol. 2. Principles of Political Economy, Part I*. Edited by John M. Robson. University of Toronto Press.
- Mill, John Stuart. 1965b. *Collected Works of John Stuart Mill: Vol. 3. Principles of Political Economy, Part II*. Edited by John M. Robson. University of Toronto Press.
- Mill, John Stuart. 1985. *Collected Works of John Stuart Mill: Vol. 10. Utilitarianism*. Edited by John M. Robson. University of Toronto Press.
- Mishel, Lawrence. 2022. "Not So Free to Contract: The Law, Philosophy, and Economics of Unequal Workplace Power." *Journal of Law and Political Economy* 3 (1).
- Nussbaum, Martha. 2000. "Women's Capabilities and Social Justice." *Journal of Human Development* 1 (2): 219–47. <https://doi.org/10.1080/713678045>.
- Piketty, Thomas. 2014. *Capital in the Twenty-First Century*. Translated by Arthur Goldhammer. Belknap.
- Ricardo, David. 1951. "On the Principles of Political Economy and Taxation." In *Works and Correspondence*, edited by Piero Sraffa and Maurice H. Dobb. Cambridge University Press.
- Robbins, Lionel. 1948. *An Essay on the Nature and Significance of Economic Science*. 2nd ed. Macmillan.
- Sen, Amartya. 1999. *Development as Freedom*. A.A. Knopf.
- Smith, Adam. 1776 [1981]a. *An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. 1*. Glasgow Edition of the Works and Correspondence of Adam Smith. Liberty Fund.
- Smith, Adam. 1776 [1981]b. *An Inquiry into the Nature and Causes of the Wealth of Nations, Vol. 2*. Glasgow Edition of the Works and Correspondence of Adam Smith. Liberty Fund.
- Viscusi, W. Kip. 1983. *Risk By Choice*. Harvard University Press.
- Weber, Max. 2003. *The Protestant Ethic and the Spirit of Capitalism*. Translated by Talcott Parsons. Dover.

ARTIFICIAL INTELLIGENCE

Values, Governance, and Policy

*Justin B. Biddle***45.1 Introduction**

Artificial intelligence (AI) has undergone rapid advancement and is now deployed widely. In industrialized societies, people interact with AI every day—it is built into our email clients, word processors, search engines, streaming services, and other technologies, and for many people, interaction with generative AI, including large language models (LLMs) and image generators, is commonplace. Some recent advancements in AI have created excitement and optimism. AI systems such as AlphaFold, which can be used in scientific research to predict protein folding, appear to be genuinely path breaking and beneficial ([Service 2020](#)). At the same time, there have been numerous highly-publicized instances of AI performing in ways that are discriminatory and untrustworthy. From chatbots that spew racist and sexist hate speech (e.g., [Vincent 2016](#)), to hiring algorithms that are biased against women (e.g., [Dastin 2018](#)), to image recognition algorithms that categorize Black people as gorillas (e.g., [Grant and Hill 2023](#)), we have seen a litany of examples confirming that AI systems can be biased and harmful.

It might be tempting to respond to cases like this by arguing that AI systems should be unbiased, in the sense of being free from human values. This chapter will explain why this cannot be. AI systems are necessarily value-laden. They might involve values that one endorses, such as respect for human autonomy and universal human rights. They might reflect values that are reprehensible, such as patriarchy and white supremacy. Or they might embed still other values. But they necessarily involve some values or other. Given this—and given the potential of AI to impact and disrupt society in significant ways—it is imperative that the values that influence AI systems are critically scrutinized and appropriately managed.

These conclusions that AI systems are value-laden, and that it is crucial that the values that influence AI are scrutinized, have been endorsed not just by philosophers but in computing and engineering fields more broadly. For example, the Institute of Electrical and Electronics Engineers (IEEE), one of the largest engineering professional societies in the world, writes:

Autonomous and intelligent systems (A/IS) research and design must be developed against the backdrop that technology is not neutral. A/IS embody values and biases that can influence important social processes like voting, policing, and banking. To ensure that A/IS benefit humanity, A/IS research and design must be underpinned by ethical and legal norms.

([IEEE 2019](#), 124)

This chapter will undertake two primary tasks. The first is to discuss some important respects in which AI systems are value-laden. While, as noted, it is broadly accepted that AI systems are value-laden, it is important to understand the ways in which decisions throughout the development and deployment process involve value judgments, if these values are to be critically scrutinized and managed appropriately. This discussion paves the way for the second main task of the chapter, which is to address the question of how AI systems can be “underpinned by ethical and legal norms,” so as to benefit humanity. There is much disagreement surrounding this question, and this chapter will not attempt to answer it definitively. Rather, it will discuss three recent approaches to AI governance, which shed light on ways that policymakers are attempting to address practical questions about the management of values in AI. This discussion will illustrate how controversial and important these questions are.

45.2 Values in AI

Science and technology are value-laden activities. The development of AI systems is no different in this respect.¹ Values impact not just the preliminary stages of research into AI (e.g., which projects are funded) and the ways in which AI is used, but throughout the AI lifecycle. This chapter will focus on the influence of values in the following areas: (1) problem framing and operationalization; (2) data; (3) modeling and validation; and (4) deployment.²

45.2.1 Values in problem framing and operationalization

In research and development (R&D), investigators must make judgments about which projects are worthy of pursuit. Additionally, in developing AI systems (and technological systems more generally), developers must make decisions about how to translate abstract, high-level problems into ones that are more specifically-formulated and operationalized. These decisions involve value judgments that have real-world effects on people (c.f., [Biddle 2018](#); [Elliott 2017](#); [Lacey 1999](#)).

Suppose, for example, that one is concerned about bias in the criminal justice system, or about student dropout rates, or about deficiencies in information access and processing. These problems might be addressed in many ways, some involving advanced technology and some not. If one develops AI systems to address them, one must translate these high-level problems (bias in the justice system, high student drop-out rates, and information deficiencies) into more specifically-formulated ones that allow for measurement.

To reduce bias in the criminal justice system, many organizations have developed algorithmic systems for assessing the risk that an “offender” will “re-offend.” Proponents of these systems argue that they can provide objective assessments of risk, which can be used to influence bail and sentencing in ways that reduce bias. There are, however, other ways of framing the problem. For example, instead of attempting to predict risk of recidivism, one could develop AI systems that examine decisions of judges, alerting them (or someone else) in case their decisions are biased. Similarly, there are many potential ways of addressing problems of high dropout rates and information deficits, even if one is restricting oneself to AI. One could develop AI systems that predict student success or failure and then use these predictions to attempt to reduce dropout rates—or one could develop AI systems that identify university practices (including financial aid packages, program requirements, types of assignments, grading schemes, and student support systems) that are associated with student retention (among other possibilities). One could develop language models that strive for “general intelligence”—or one could develop bespoke models tailored to specific information contexts (among other possibilities).

The question of *which problem an AI system should address* is one that cannot be answered in a value-free manner. In the cases discussed above, different problem framings locate the object of

analysis differently. For some problem framings, the object of analysis is purported offenders or students; for others, it is judges or university practices. The consequences of errors or mistakes by the AI system are very different, depending on the different problem framings. For example, the consequences of errors by recidivism prediction systems might be that a purported offender is unjustly denied bail or sentenced to prison, whereas the consequences of error for systems that assess decisions of judges might be that a judge is unjustly denied a career-advancement opportunity. Decisions about problem framing reflect judgments about which mistakes one is more or less willing to tolerate—as well as judgments about which groups of people one is more or less willing to expose to risk. These judgments reflect values.

While some might think that questions about problem framing are “external” to AI development, there are other problem-framing decisions—relating to the operationalization of concepts employed within AI systems—that pertain to validity. These decisions are also value-laden. For example, a recidivism-prediction system (and a student-success algorithm) requires that the concept of recidivism (and student success) be operationalized. Recidivism might be operationalized in many ways—but most developers operationalize it in terms of arrest. That is, someone is classified as a recidivist if they are arrested and then arrested again (within a specified period of time). If an AI system operationalizes recidivism in this way, it is not, strictly speaking, assessing the risk that someone who has committed a crime will commit another crime; it is assessing the risk that someone who has been arrested will be arrested again. Operationalizing in terms of arrest arguably reduces construct validity, and it exposes to harm populations that are disproportionately subject to arrest, including false arrest.

Similarly, student success might be operationalized in terms of a combination of grade point average (GPA) upon graduation, post-graduate employment, service to the community, reports of mental health and life satisfaction, and other factors—or it could be operationalized in terms of GPA alone. AI systems that operationalize student success in terms of GPA are likely to be biased in favor of students who come from privileged backgrounds with strong academic and/or social support systems, or who have the financial resources to attend university without working jobs to repay loans, or who enroll in classes with grade inflation. AI applications involve theoretical constructs that require operationalization, and how they are operationalized reflects value judgments about which groups may be impacted, and how (Biddle 2022; Fazelpour and Danks 2021; Jacobs and Wallach 2021).

45.2.2 Values in data

In the case of AI systems developed with machine learning, performance will depend on the data used to train them. Developers must make judgments about the quantity of data used to train the system, the quality of these data (e.g., how accurate the data are with respect to desired attributes), and/or the diversity of data (e.g., how representative the data are relative to groups the system is likely to be used on), among other things. In some cases, they must make decisions about whether it is fair to use a data set for training purposes, or whether use would violate copyright or other rights. These decisions about training data reflect value judgments in important ways.

Many data sets reflect historical injustices and biases, and if AI systems are trained on biased data, then their performance will likely be biased. Perhaps the most well-known example of biased data leading to biased predictions concerns facial recognition. AI systems that were trained to recognize faces were trained primarily on images of lighter skinned people (especially men). As a result, while the systems performed well at identifying lighter skinned individuals, they had disproportionately high error rates for darker skinned individuals, especially darker skinned women (e.g., Buolamwini and Gebru 2018). There are many other real-world examples of biased training data leading to biased predictions. Some algorithms that make hiring recommendations have been found to be biased against women, in part because they were trained on data that reflect historical hiring decisions that

advantaged men over women (e.g., [Dastin 2018](#)). Algorithms used in policing in the United States have been found to be biased against Black and brown people, in part because the algorithms are trained on arrest data, and Black and brown individuals are disproportionately subject to arrest, relative to white people (e.g., [Angwin et al. 2016](#); [Biddle 2022](#), 2023b). This phenomenon of biased training data leading to biased performance can generate feedback loops of injustice—AI systems that are trained on data reflecting historical injustices can lead to predictions that further disadvantage groups that are already historically disadvantaged.

It might be tempting to respond to these concerns by urging AI developers to ensure that training data are unbiased. There is some merit to this response, as biases in data sets can often be reduced by improving data curation practices, ensuring greater diversity in data sets, and the like ([Gebru et al. 2021](#)). But it is not possible to create data sets in a completely value-neutral way. Developers must make judgments about whether existing data sets are “good enough” for purposes of training AI—and if not, which steps to take, and how many resources to expend, to improve the data. They also must, in many cases, make decisions about how to structure data. These decisions inescapably involve value judgment.³

As an illustration, consider the variety of value-laden judgments that Joy Buolamwini and Timnit Gebru made during the course of their pioneering study of commercial facial recognition systems (2018). Prior to their study, facial recognition systems were advertised as being highly accurate, but there was little data that analyzed accuracy rates across different intersectional categories (such as race and gender). Buolamwini and Gebru were concerned that the systems worked much better on lighter skinned individuals than darker skinned ones—based partly on Buolamwini’s personal experience of not being recognized by them. In designing their study, they needed to decide whether to use existing data sets as benchmarks or to create a new one. Given that existing data sets contained disproportionate numbers of lighter skinned faces, they decided to create a new one—a decision that prioritized diversity and representation over other values, such as convenience. Once they made this decision, they needed to decide how the data set should be structured—for example, whether individuals in the data set should be categorized by race or ethnicity, or by some other factors, and whether individuals would be represented in terms of binary sex categories. They decided to categorize individuals in terms of skin types rather than race or ethnicity. Doing this allowed them to make distinctions between lighter or darker skinned individuals who identified as the same race. Additionally, they chose to represent individuals in terms of binary sex categories, despite believing that this categorization “does not adequately capture the complexities of gender or address transgender identities” ([Buolamwini and Gebru 2018](#), 6). They did this because they were attempting to assess commercial gender classification systems, which categorized individuals in terms of binary sexes. These decisions about what counts as “good enough” data and about how data should be structured involve value judgments about which types of harms, and for which populations, are more or less tolerable, and they reflect, reinforce, or critique dominant representations and categorizations of people. How these judgments are made can impact the evaluation of new technologies, which in turn significantly impact different populations.

In addition to judgments about data quantity, quality, and diversity, developers must in many cases make value-laden judgments about whether they can fairly access data for training purposes. For example, after Facebook developed its facial recognition tools, it was criticized for using data from users who had not consented to this use (e.g., [Singer 2018](#)). Developers of generative AI applications are grappling with questions about the conditions under which they may fairly use data (such as text, visual art, or music) to train systems that will generate content. Some have argued that the use of copyright-protected art, text, or music to train generative AI systems amounts to theft (e.g., [Goetze 2024](#)). Whether or not that is the case, judgments about fair data use have significant societal implications—including for the ability of writers, artists, musicians, and others to make a living. Judgments by developers about the conditions under which they may access data for training purposes involve values.

45.2.3 Values in modeling and validation

Given a problem framing and data set, developers can construct models that perform decision-making tasks such as classification or generation. They employ learning algorithms to model some portion of the data (the “training set”), and they test the performance of the model based on how well it predicts portions of the data that were not used to train it (the “test set”). The process of constructing and validating a model is iterative, and developers must make judgments about the criteria of success according to which their models are evaluated. This is a stage in the AI development process in which highly technical decisions can encode values that are both consequential and difficult to spot.⁴

One important decision at this stage that involves values is the choice of an objective function. Developers might, for example, decide to optimize for overall accuracy. Alternatively, they might optimize in ways that account for disparities in accuracy across different groups. One of the facial recognition systems examined by [Buolamwini and Gebru \(2018\)](#) had an overall accuracy rate of 93.7%—but it also had significant disparities in accuracy across groups (100% for lighter skinned males and 79.2% for darker skinned females). The decision to optimize for overall accuracy can mask performance disparities between groups and can disproportionately harm some groups. Additionally, optimizing for overall accuracy can mask disparities in the *types of errors* across different groups. In many cases, the impacts of Type I errors (false positives) and Type II errors (false negatives) are very different. In the case of recidivism prediction systems, the error of predicting that someone will reoffend when they don’t can have very different consequences than predicting that they will not reoffend when they do (e.g., receiving a punishment that is more stringent or more lenient, respectively, than is merited) (e.g., [Angwin et al. 2016](#); [Biddle 2022](#)). Similarly, in the case of student success algorithms, different types of errors could lead to disparities in opportunities for different students (e.g., [Fazelpour and Danks 2021](#)). In many cases, decisions about optimization and constraints will inescapably involve values—different decisions will involve tradeoffs between different goods, and whatever decision is made will reflect the prioritization of some goods over others. Constraining acceptable solutions to ones that satisfy some fairness metric might involve an escapable tradeoff between overall accuracy and that conception of fairness. Furthermore, different fairness metrics are, in many cases, incompatible with one another (e.g., [Chouldechova 2017](#)). In cases in which there are tradeoffs between accuracy and a fairness metric, or tradeoffs between different fairness metrics, decisions about optimization and constraints will involve value judgments about which of these aims should be prioritized and which may be sacrificed.

In creating AI chatbots to interact with users, developers must make morally-significant decisions about the constraints that should guide what a language model may and may not “say.” Microsoft’s Tay is an infamous example of what can happen if little attention is paid to such guardrails. What happened in this episode is succinctly described by the title of an article from *The Verge*: “Twitter Taught Microsoft’s AI Chatbot to be a Racist Asshole in Less than a Day” ([Vincent 2016](#)). Developers have learned from this debacle, and there is now much attention paid to constraints on the appropriateness of chatbot outputs. But it is clear that decisions about guardrails—whether there should be any, and if so, what they should be—involve morally significant judgments. What counts as offensive speech? What kind of speech should be off-limits? How should chatbots respond to queries about controversial or highly-polarized topics? There is no way of answering these questions in a value-neutral manner.

The decisions discussed above involve value judgments about the relative prioritization of success metrics. Additionally, developers must make value-laden decisions about how successful an AI system needs to be—including how much testing a model should undergo—before it is deployed. Facebook’s original motto was “move fast and break things.” This motto, extended into the realm of AI, might lead developers to push AI systems onto the market quickly, before they undergo extensive testing—and to try to clean up whatever damage is done after it occurs. Others have argued

that developers should take a precautionary approach, doing extensive safety testing before putting new AI systems onto the market. These debates illustrate conflicting views on how much testing and validation a model should undergo before it is deployed—as well as conflicting value judgments about freedom to conduct business, protection of publics from risky technologies, and others.⁵

45.2.4 Values in deployment

Most AI systems are developed in order to be put out into the world and used by organizations and individuals. This section discusses briefly some respects in which values are implicated in deployment by organizations, and it highlights how decisions at earlier stages in the R&D process—including problem framing and operationalization, data, and modeling and validation—have implications for responsible deployment.

When organizations deploy AI, they necessarily make value-laden decisions. Especially important in this context are decisions about transparency, human involvement, and accountability.⁶ Suppose, for example, that a university deploys AI to determine whether its researchers have ties to sanctioned organizations—and then uses these results to fire employees (c.f., [Saul 2025](#)). Among the many important questions we might ask about this deployment are as follows: Has the university been transparent about its use of AI? To what extent are there human beings involved (or “in the loop”) of these decision-making processes? For example, if an AI finds a potential link between a researcher and a terrorist organization, are there humans who carefully assess whether this link is real? If a university fires a researcher based on an AI output, does this researcher have recourse available to them? Can they demand an explanation of how the AI came to this conclusion or a review of the decision by human beings? In case the AI makes a mistake, who is accountable, and in what ways? These deployment decisions are highly significant, and they are often made behind closed doors and out of sight of those who are potentially impacted by AI. If AI is to be deployed responsibly, then those who are impacted by AI decisions should understand and ideally have input into how these decisions are made.

In addition to addressing the issues of transparency, human involvement, and accountability, responsible deployment of AI requires appropriate alignment between the context of development and the problems that it is deployed to address. Recall the examples of recidivism prediction and student success prediction. If recidivism is operationalized in terms of arrest, then a recidivism-prediction system does not, strictly speaking, predict the risk that someone who has committed a crime will commit another crime. It is predicting whether someone who has been arrested will be arrested again. To the extent that arrest is an inaccurate proxy for crime that reflects biases in policing priorities, deploying the system in bail or sentencing contexts could exacerbate existing inequalities and injustices. Similarly, if a student success algorithm operationalizes student success solely in terms of GPA, then deploying it in ways that tend to distribute opportunities and resources to those likely to be “successful”—as they might be used in admissions departments—could unjustly disadvantage those who are already disadvantaged.

There should also be appropriate alignment between the data used to train a system and the population context in which it will be deployed. Data about student success in elite universities with large endowments—however student success is operationalized—are unlikely to be representative of underfunded minority-serving institutions, or of colleges or universities in other countries. Similarly, data about “offenders” from one locality (e.g., Broward County Florida, from which data for the COMPAS algorithm was collected) is unlikely to be representative of other localities with different populations ([Angwin et al. 2016](#)). If there are significant differences between the population context used to train an AI system and that in which it is deployed, there will be significant risk of external invalidity—or risk that the inferences drawn by the system (e.g., prediction and classification) will be invalid when made outside of the population context in which the system was trained.

45.3 AI governance and policy

Thus far, this chapter has examined the ways in which AI systems are value-laden. The value-laden character of AI systems raises further questions about how these values should be managed. Questions about which values should influence science, and how they should do so, have been addressed extensively by philosophers of science (e.g., Douglas 2009; Elliott 2022; Hicks 2014; Holman and Wilholt 2022). In the case of AI, however, it is not only academics who have addressed these questions, but also professional organizations, industry, and governments. Organizations and governments around the world are developing structures and frameworks of various kinds that purport to influence how AI systems are shaped.⁷

This section examines three competing policy and governance frameworks for AI, one from the European Union (EU) and two from the United States (US). Among other aims, these frameworks all attempt to facilitate what philosophers have termed the “management of values” in science and technology—though in very different ways. More specifically, they attempt to influence *which values* guide the development and deployment of AI, and the *systemic structure* of AI governance.⁸ These approaches—which are far from exhaustive—indicate the practical relevance of philosophical questions about the management of values in science and technology. They also indicate how controversial these issues have become—and how important they are. First, this section will discuss the EU AI Act, which establishes mandatory requirements for incorporating values into AI. This will be followed by a discussion of the *Artificial Intelligence Risk Management Framework*, created by the US National Institute of Standards and Technology (NIST), which is a voluntary framework for facilitating trustworthy AI. The chapter will conclude with discussion of policy proposals by the second Trump Administration and advisors to it, which involve a very different approach to managing values in AI—one that rejects values such as diversity, equity, and inclusion.

The EU AI Act aims to promote *human-centric* and *trustworthy* AI—that is, AI that supports human autonomy and is worthy of the trust of all human beings. As a part of the effort to promote trustworthy AI, the Act attempts to ensure “a high level of protection of health, safety, fundamental rights enshrined in the Charter of Fundamental Rights, including democracy, the rule of law and environmental protection, against the harmful effects of artificial intelligence systems” (Chapter 1, Article 1). To do this, the Act divides AI systems into five risk categories, and it imposes legally binding requirements on developers and deployers of AI, corresponding to these risk categories.

Under this framework, some AI systems are classified as posing “unacceptable risk” to individuals; they cannot be deployed without violating fundamental rights, and as such, they are banned outright. These include AI systems for general social scoring, for unconsciously manipulating individuals in ways that could lead to significant and direct harm, and systems for remote biometric identification in public. The next category is “high-risk systems.” These are systems that are legally permitted, but they pose risks that are sufficiently significant that they must undergo heightened regulatory oversight. High-risk systems include those used in critical infrastructures, education, employment, law enforcement, migration, asylum, border control, and others. Many of the controversial AI systems that have been discussed in this chapter—including recidivism-prediction systems, student success algorithms, and AI-based hiring recommendations—fall under this category. Under the Act, high-risk systems will (again) be subject to heightened regulatory oversight, including requirements of pre-market approval and post-market monitoring. The lower three risk categories involve (for the most part) lesser requirements and will not be discussed here.

Which values should shape the development and deployment of AI? The reference to the Charter of Fundamental Rights of the European Union helps to clarify this. The Charter is grounded in the “universal values of human dignity, freedom, equality and solidarity,” and it includes 54 Articles, many of which are explicitly highlighted in the EU AI Act. These include the right to human dignity (Article 1); protection of personal data (Article 8); freedom of expression and information

(Article 11); right to education (Article 14); non-discrimination (Article 21); gender equality (Article 23); rights of persons with disabilities (Article 26); workers' rights (Articles 27–31), and many others. Within the Act, the “universal human values of human dignity, freedom, equality and solidarity”—and the rights and protections that fall underneath them—are understood as political values that are necessary for individuals to be able to choose lives that are good for them. Overall, the Act attempts to manage values in AI by establishing a regulatory system intended to implement political and legal guardrails based on fundamental human rights (c.f., [Schroeder 2020](#)).

A second approach to managing values in AI is represented by NIST's *AI Risk Management Framework* ([Tabassi 2023](#)). Like the EU AI Act, the NIST AI RMF emphasizes the concept of *trustworthy AI*, which it defines in terms of values that AI systems should reflect. These values are safety, fairness, privacy, accountability and transparency, explainability and interpretability, security and resilience, and validity and reliability ([Tabassi 2023](#), 12). Also like the EU AI Act, NIST emphasizes values that should (on its view) be acceptable to all. Unlike the EU AI Act, the NIST AI RMF does not impose any mandatory requirements on developers or deployers of AI. Rather, it is intended to serve as a resource to manage, on a voluntary basis, “the many risks of AI and promote the trustworthy and responsible development and use of AI systems” ([Tabassi 2023](#), 2). In order to promote trustworthy AI, it recommends that organizations develop cultures and processes for governing, mapping, measuring, and managing AI risks ([Tabassi 2023](#), 12).

Even though the NIST AI RMF is a voluntary framework, it still played an important role in AI policy in the US under the Biden Administration. It was referenced many times in the White House Executive Order on Safe, Secure, and Trustworthy AI, an expansive Executive Order that articulated a vision for developing and deploying AI responsibly, in ways that safeguard civil rights—including the rights of historically marginalized groups—and promote international cooperation ([Executive Order No. 14110 2023](#)). Among other things, the Executive Order required the Secretary of Commerce to establish voluntary guidelines and best practices for industry for developing and deploying trustworthy AI. Additionally, it placed mandatory requirements on the development and deployment of AI by federal agencies, including “minimum risk-management practices for Government uses of AI that impact people's rights or safety” and “assessing and mitigating disparate impacts and algorithmic discrimination” ([Executive Order No. 14110 2023](#), 75218). The NIST AI RMF was to serve as a building block in these endeavors and as an important resource for managing values in AI.

While the NIST AI RMF and the EU AI Act differ in their approaches to regulation, they both emphasize the importance of facilitating the management of AI in accordance with values that many endorse, including respect for human dignity and autonomy, safety, privacy, transparency and explainability, accountability, and fairness and justice. Moreover, it is not just NIST and the EU that have highlighted these values in the context of AI. Ethics and governance documents published by many different organizations around the world—including in industry, government, and professional organizations—have emphasized the importance of developing and deploying AI in accordance with these values, and some meta-analyses of these documents have argued that there is an emerging convergence around these values (c.f., [Floridi and Cows 2019](#); [Jobin, Ienca, and Vayena 2019](#); [Schiff et al. 2021](#)). Such a convergence, however, was illusory, as recent events have made clear.

Soon after beginning his second term as President, Donald Trump signed an executive order entitled, “Removing Barriers to American Leadership in Artificial Intelligence,” which revoked Biden's White House Executive Order on Safe, Secure, and Trustworthy AI ([Executive Order No. 14179 2025](#)). Trump's executive order, which sets as policy to “sustain and enhance America's global AI dominance,” states that “we must develop AI systems that are free from ideological bias or engineered social agendas” ([Executive Order No. 14179 2025](#), 8741). While the executive order does not explain what freedom from “ideological bias” and “engineered social agendas” mean, other activities by the Trump Administration make clear that many of the values highlighted in the NIST AI RMF—and

in numerous other AI governance frameworks—are repudiated. For example, while the NIST AI RMF highlights workforce diversity, equity, inclusion, and accessibility as important for facilitating trustworthy AI (Tabassi 2023, 23), the Trump Administration has issued executive orders that combat “dangerous, demeaning, and immoral” policies based on these values (e.g., [Executive Order No. 14173 2025](#), 8633). It has also issued orders to remove “woke” words and concepts from the policies and initiatives of the federal government. These include variations of words such as “diversity,” “equity,” “inclusion,” “accessible,” “bias,” “discrimination,” “clean energy,” “climate science,” “gender,” and “racial” ([Yourish et al. 2025](#)). Variations of each of these terms were used either in the NIST AI RMF or the 2023 Executive Order. Both emphasized fairness in AI and the mitigation of risks of discrimination, and the 2023 executive order emphasized the role that AI could play in fostering clean energy and facilitating climate resilience.

As this chapter has argued, there is no such thing as AI that is value-free. AI systems necessarily reflect values, including values that are ethically and/or politically valenced. In attempting to purge particular values from AI, such as particular conceptions of fairness and justice, Trump’s executive order does not facilitate AI that is free from an engineered social agenda—it merely facilitates a different engineered social agenda, one in which efforts to advance diversity, equity, inclusion, and accessibility are seen as “dangerous, demeaning, and immoral.”

An elaboration of the engineered social agenda that underpins aspects of Trump’s AI policy can be found in “The Techno-Optimist Manifesto,” written by Marc Andreessen, a venture-capitalist and influential technology advisor to Trump. After expressing his faith that technological innovation fueled by unrestrained market forces will lead to social progress, he characterizes the “enemies” of progress and innovation.

We have enemies. Our enemies are not bad people—but rather bad ideas. Our present society has been subjected to a mass demoralization campaign for six decades—against technology and against life—under varying names like “existential risk”, “sustainability”, “ESG”, “Sustainable Development Goals”, “social responsibility”, “stakeholder capitalism”, “Precautionary Principle”, “trust and safety”, “tech ethics”, “risk management”, “de-growth”, “the limits of growth”. This demoralization campaign is based on bad ideas of the past—zombie ideas, many derived from Communism, disastrous then and now—that have refused to die.

(Andreessen 2023)

This manifesto pushes a social agenda, one based on the elimination of guardrails or protections. Many of these “zombie ideas”—including sustainability, trust and safety, social responsibility, ethics, precaution, and risk management—constitute guardrails on AI that arguably promote welfare for all, not just for a few. The assertion that these guardrails should be removed is an essential aspect of the “techno-optimist” social agenda. On this approach, values in AI will still be managed. Which values, then, will shape the development and deployment of AI? The values of those who have the power and resources to structure markets, including venture capitalists (like Andreessen) who influence which technological projects are funded and which are not. A discussion of whether a “techno-optimist” social agenda promotes human welfare for all, rather for a select few, is beyond the scope of this chapter. But we should not be misled into thinking that it does not constitute an “engineered social agenda.”

45.4 Conclusion

AI systems are necessarily value-laden. Which values influence AI, however, is a matter of social, political, and economic contestation. Many in academia—including many philosophers of science, computer scientists and engineers, and others—hold that values such as respect for human dignity

and autonomy, safety, privacy, transparency and explainability, accountability, and fairness and justice should govern the development and deployment of AI. Some of these values, however, are currently under attack, including by many who are prominent in government. Given the increasing importance of AI and its ability to disrupt society in significant ways, current and future decisions about how values in AI should be managed will be enormously consequential for humanity.

Notes

- 1 This chapter will follow the US National Institute of Standards and Technology (NIST) in defining “AI system” as “an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments” (Tabassi 2023, 1).
- 2 In dividing the influence of values into these categories, I draw on Fazelpour and Danks (2021), as well as Biddle (2022, 2023a). These categories should not be taken to be exhaustive.
- 3 For discussion of tradeoffs and data diversity in the biomedical sciences, see Nieves Delgado et al. (2026).
- 4 For discussions of values and models in science more generally, see Longino (2002) and Harvard and Winsberg (2022).
- 5 Decisions about how much testing and validation models should undergo before they are deployed relate in important ways to decisions about inductive risk management. On this issue, see Douglas (2009), Elliott and Richards (2017), Havstad (2022), Rudner (1953), and many others.
- 6 This list is not exhaustive. For further discussion of these and related issues, see European Commission (2019).
- 7 Many AI firms have created organizational structures—including ethics offices, advisory committees, and others—that purportedly facilitate the management of values in AI. Examination of these structures is outside of the scope of this chapter. See Biddle, Nelson, and Olugbade (2024) and Schuett, Reuel, and Carlier (2025) for further discussion.
- 8 Holman and Wilholt (2022) have articulated five types of “demarcation strategies” for distinguishing between appropriate and inappropriate value influences in science: axiological, functionalist, consequentialist, coordinative, and systemic. Prominent AI governance frameworks, including the ones discussed here, focus (at least) on axiological, consequentialist, coordinative, and systemic strategies. I omit discussion of consequentialist and coordinative strategies due to limitations of space.

References

- Andreessen, Marc. 2023. “The Techno-Optimist Manifesto.” <https://a16z.com/the-techno-optimist-manifesto/>.
- Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. “Machine Bias.” *ProPublica*, May 23. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.
- Biddle, Justin. 2018. “‘Antiscience Zealotry’? Values, Epistemic Risk, and the GMO Debate.” *Philosophy of Science* 85: 360–79. <https://doi.org/10.1086/697749>.
- Biddle, Justin. 2022. “On Predicting Recidivism: Epistemic Risk, Tradeoffs, and Values in Machine Learning.” *Canadian Journal of Philosophy* 52 (3): 321–41. <http://doi.org/10.1017/can.2020.27>.
- Biddle, Justin. 2023a. “Values in AI Systems.” In *Routledge Companion to Technology Ethics*, edited by Gregory Robson and Jonathan Tsou. Routledge.
- Biddle, Justin. 2023b. “Organizations and Values in Science and Technology.” *Philosophy of Science* 91 (5): 1159–68. <https://doi.org/10.1017/psa.2023.145>.
- Biddle, Justin, John P. Nelson, and Olajide Olugbade. 2024. “How Can We Know if You’re Serious? Ethics Washing, Symbolic Ethics Offices, and the Responsible Design of AI Systems.” *Canadian Journal of Philosophy* 54(4): 329–45. <https://doi.org/10.1017/can.2025.9>.
- Buolamwini, Joy, and Timnit Gebru. 2018. “Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.” *Proceedings of Machine Learning Research* 81: 77–91.
- Chouldechova, Alexandra. 2017. “Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments.” *Big Data* 5 (2). <https://doi.org/10.1089/big.2016.0047>.
- Dastin, Jeffrey. 2018. “Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women.” *Reuters*, October 10. <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/>.

- Douglas, Heather. 2009. *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt6wrc78>.
- Elliott, Kevin. 2017. *A Tapestry of Values*. Oxford University Press.
- Elliott, Kevin. 2022. *Values in Science*. Cambridge University Press. <https://doi.org/10.1017/9781009052597>.
- Elliott, Kevin, and Ted Richards, eds. 2017. *Exploring Inductive Risk: Case Studies of Values in Science*. Oxford University Press.
- European Commission, High Level Expert Group on AI. 2019. “Ethics Guidelines for Trustworthy AI.” <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.
- Executive Order No. 14110. 2023. 88 FR 75191. <https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>.
- Executive Order No. 14173. 2025. 90 FR 8633. <https://www.federalregister.gov/documents/2025/01/31/2025-02097/ending-illegal-discrimination-and-restoring-merit-based-opportunity>.
- Executive Order No. 14179. 2025. 90 FR 8741. <https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>.
- Fazelpour, Sina, and David Danks. 2021. “Algorithmic Bias: Senses, Sources, Solutions.” *Philosophy Compass* 16: e12760. <https://doi.org/10.1111/phc3.12760>.
- Floridi, Luciano, and Josh Cows. 2019. “A Unified Framework of Five Principles for AI in Society.” *Harvard Data Science Review* 1 (1). <https://doi.org/10.1162/99608f92.8cd550d1>.
- Gebru, Timnit, Jaime Morgenstern, Briana Vecchione, et al. 2021. “Datasheets for Datasets.” *Communications of the ACM* 64(12): 86–92. <https://doi.org/10.1145/3458723>.
- Goetze, Trystan. 2024. “AI Art is Theft: Labour, Extraction, and Exploitation: Or, On the Dangers of Stochastic Pollocks.” In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24)*. Association for Computing Machinery. <https://doi.org/10.1145/3630106.3658898>.
- Grant, Nico, and Kashmir Hill. 2023. “Google’s Photo App Still Can’t Find Gorillas. And Neither Can Apples.” *New York Times*, May 22. <https://www.nytimes.com/2023/05/22/technology/ai-photo-labels-google-apple.html>.
- Harvard, Stephanie, and Eric Winsberg. 2022. “The Epistemic Risk in Representation.” *Kennedy Institute of Ethics Journal* 32 (1): 1–31. <https://doi.org/10.1353/ken.2022.0001>.
- Havstad, Joyce. 2022. “Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk.” *Canadian Journal of Philosophy* 52 (3): 295–320. <https://doi.org/10.1017/can.2021.15>.
- Hicks, Dan. 2014. “A New Direction for Science and Values.” *Synthese* 191 (14): 3271–95. <https://doi.org/10.1007/s11229-014-0447-9>.
- Holman, Bennett, and Torsten Wilholt. 2022. “The New Demarcation Problem.” *Studies in History and Philosophy of Science* 91: 211–20. <https://doi.org/10.1016/j.shpsa.2021.11.011>.
- IEEE. 2019. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*, First Edition. IEEE. <https://www.ethics.org/wp-content/uploads/Ethically-Aligned-Design-May-2019.pdf>.
- Jacobs, Abigail, and Hanna Wallach. 2021. “Measurement and Fairness.” In *Conference on Fairness, Accountability, and Transparency (FAccT '21)*. Association for Computing Machinery. <https://doi.org/10.1145/3442188.3445901>.
- Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. “The Global Landscape of AI Ethics Guidelines.” *Nature Machine Intelligence* 1: 389–99. <https://doi.org/10.1038/s42256-019-0088-2>.
- Lacey, Hugh. 1999. *Is Science Value Free? Values and Scientific Understanding*. Routledge.
- Longino, Helen. 2002. *The Fate of Knowledge*. Princeton University Press.
- Nieves Delgado, Abigail, Jan Baedke, and Ann-Christin Fischer. 2026. “The Limits of Diversity in Science: The Case of Human Microbiome Research.” In *The Routledge Handbook of Values and Science*, edited by Kevin C. Elliott and Ted Richards. Routledge.
- Rudner, Richard. 1953. “The Scientist Qua Scientist Makes Value Judgements.” *Philosophy of Science* 20 (1): 1–6. <https://doi.org/10.1086/287231>.
- Saul, Stephanie. 2025. “Yale Suspends Scholar After A.I.-Powered News Site Accuses Her of Terrorist Link.” *New York Times*, March 12. <https://www.nytimes.com/2025/03/12/us/yale-suspends-scholar-terrorism.html>.
- Schiff, Daniel, Jason Borenstein, Justin Biddle, and Kelly Laas. 2021. “AI Ethics in the Public, Private, and NGO Sectors: A Review of a Global Document Collection.” *IEEE Transactions on Technology and Society* 2 (1): 31–42. <https://doi.org/10.1109/TTS.2021.3052127>.
- Schroeder, S. Andrew. 2020. “Thinking about Values in Science: Ethical vs. Political Approaches.” *Canadian Journal of Philosophy* 52 (3): 246–55. <https://doi.org/10.1017/can.2020.41>.

- Schuett, Jonas, Ann-Katrin Reuel, and Alexis Carlier. 2025. "How to Design an AI Ethics Board." *AI and Ethics* 5: 863–81. <https://doi.org/10.1007/s43681-023-00409-y>.
- Service, Robert. 2020. "'The Game has Changed.' AI Triumphs at Protein Folding." *Science*, December 4. <https://www.science.org/doi/10.1126/science.370.6521.1144>.
- Singer, Natasha. 2018. "Facebook's Push for Facial Recognition Prompts Privacy Alarms." *New York Times*, July 9. <https://www.nytimes.com/2018/07/09/technology/facebook-facial-recognition-privacy.html>.
- Tabassi, Elham. 2023. "Artificial Intelligence Risk Management Framework (AI RMF 1.0)." NIST Trustworthy and Responsible AI, National Institute of Standards and Technology, Gaithersburg, MD, [online]. https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=936225
- Vincent, James. 2016. "Twitter Taught Microsoft's AI Chatbot to be a Racist Asshole in Less than a Day." *The Verge*, March 24. <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>.
- Yourish, Karen, Annie Daniel, Saurabh Datar, Isaac White, and Lazaro Gamio. 2025. "These Words Are Disappearing in the New Trump Administration." *New York Times*, March 7. <https://www.nytimes.com/interactive/2025/03/07/us/trump-federal-agencies-websites-words-dei.html>.

INDEX

Note: Page references in **bold** denote tables and with “n” endnotes.

- abolitionism 271–274; algorithmic 269–274; Fryer on 272; vs. gradualism 271–274; police 272
- absolute conception 43
- academic migration 156
- academic publishing 156–157
- accountability 282, 286, 288–289, 447–448
- achievement thesis 38–39
- activism 38, 193, 329
- Adams, Mark B. 245
- advocating for research on sex differences 457–459
- aesthetics of science 89–91
- aesthetic values 49; overview 84–85; in scientific practice 84–91; systemic bias 87–88; *see also* perceptual beauty
- agenda setting 295–296
- agenda-setting power 343–345
- agnotology 296; *see also* ignorance
- agricultural genomics research 326–333
- agriculture 169, 243, 257, 326, 328, 348, 375, 419, 421
- AI chatbots 515
- ‘Āina Back Summit program 349, 359n1
- Albuquerque Declaration 356
- Alcoff, Linda M. 168
- Alexandrova, Anna 54, 204
- algorithmic abolitionism 269–274; abolitionist horizons 276; and anti-ethics 269–274; and ethics 269–274; fairness metrics and modular reform 264–265; incrementalism 264–265; non-ideal theory of scientific legitimacy 274–276; and Racial Algorithm 263–276; Racial Contract 266–269
- algorithmic training 268
- Allali, Imane 166
- All of Us* cohort study 471, 473–475
- Al-Qaeda 375
- Alzheimer’s disease (AD) 213, 458
- Ambrosj, Jacopo 448–449
- American Academy for the Advancement of Science 304
- American health libertarianism 368
- amplifiers of evidence 340–342
- Andreessen, Marc 519
- androcentrism 457; advocating for research on sex differences 457–459; battling 457–459
- Anscomb, Claire 90
- ante-ethics 271
- anthrax 375, 381
- anthropology 165, 170, 467
- anti-ethics 269–274; vs. insurrectionist ethics 270–271
- Appeal* (Walker) 270
- Aratus* 419–420
- Arcangeli, Margherita 89
- argument from inductive risk (AIR) 4, 53–56, 64; *see also* inductive risk
- arguments: aims 56–57; blurred boundaries 57–58; conceptual argument 54–55; contingency 55–56; gap argument 53; from inductive risk 53–54, 403–404, 450–451; inevitability 78–80; from plurality of epistemic values 450; against value-free ideal 52–59; from value-laden background assumptions 449
- Aristotle 260n1
- arrogance 154, 155
- artificial intelligence (AI) 12, 84, 148, 263, 269, 274, 511–520; AI governance and policy 517–519; values in AI 512–516; values in data 513–514; values in deployment 516; values in modeling and

- validation 515–516; values in problem framing/
operationalization 512–513
- Artificial Intelligence Risk Management Framework
517–519
- assessments: conceptualization 342–343; *see also*
climate assessment reports
- Attwell, Katie 367–368
- Australian National University (ANU) 331
- autonomy 282–283
- axiological demarcation strategies 305; *see also*
demarcation strategies
- bad science 27
- Baker, Rachel 200
- Ballung* concept 46
- Banal-Estañol, Albert 298
- Barber, Bernard 220
- Baric, Ralph 377
- Bayesianism (or Bayesian reasoning or analysis) 65;
see also personal probabilities
- beauty (*see* perceptual beauty)
- Beauty and Revolution in Science* (McAllister) 87
- Beauty and Truth* (Chandrasekhar) 86
- behavior beyond reproduction 461–462
- Bergen, Candice 363
- Bernal, J. D. 223
- Betz, Gregor 65, 131, 179, 484–485
- Bhagwan Shree Rajneesh 375
- bias; *see also* systemic bias
- Biddle, Justin 53, 328
- Biden, Joe 374, 379–380, 483, 518
- bifurcation point 136–137; and strategy/ies 137–139
- big data 162, 166
- Bikaraan-Behesht, Hamed 68
- Biological and Toxin Weapons Convention 374
- biology 24, 27, 55, 86, 88, 191, 231, 243, 329, 464;
conservation 489, 491, 496–497; molecular 329,
331; *see also* agricultural genomics research,
biomedical research, conservation biology,
genomics, human genomics, genome-wide
association studies, human microbiome research,
molecular biology, “sex as a biological variable”
(SABV)
- biomedical ethics 13; *see also* environmental ethics,
research ethics
- biomedical research: dual use 374–383; sex
essentialism in 457–464; *see also* biology, dual
use biomedical research, genomics, human
genomics, genome-wide association studies,
human microbiome research, molecular biology,
“sex as a biological variable” (SABV)
- Black Lives Matter 266
- Blair, Tony 369
- Bliss, Catherine 164
- blurred boundaries argument 57–58
- Bolshevik revolutionaries 242
- Branch, T. Y. 196n8, 436
- Braverman, Harry 506
- Brown, Matthew J. 25, 29, 47, 54–55, 58–59, 131,
134, 431
- Brundtland Report* 417
- Bruner, Justin P. 58, 317–319, 322
- Cabrera, Frank 114
- Cairney, Paul 370
- Canadian Charter of Rights and Freedoms (the
“Charter”) 371n5
- cancel culture 307, 312n1
- capabilities approach to human welfare 507–509
- Carino, Joji 352, 355
- Carson, Rachel 306
- Cartwright, Nancy 46
- Chandrasekhar, Subrahmanyan 86–87
- chemistry 12, 86, 295
- Chesterton, G.K. 112–113
- ChoGlueck, Chris 114
- Chouldechova, Alexandra 264, 266
- Churchman, C. West 53
- citizen science 202, 414, 438–440; *see also*
community-based participatory research,
participant-led science, public participation
model, stakeholder involvement
- Citu, Florin 369
- civic epistemology 233, 235–236
- classical economics: contrasting values in
501–502; work ethic values influenced
503–505
- classification 7, 28, 168, 191, 264, 420, 438,
468–471, 514–516
- climate assessment reports 339–345; agenda-setting
power 343–345; amplifiers of evidence 340–342;
attention cultivation 343–345; conceptualization/
synthesis orientation of assessment 342–343;
diagrams 340–342; epistemic curation 343–345;
error and silence 339–340; graphs 340–342; visual
representation 340–342; *see also* assessment
- climate change: institutionalization and political
constitution 338–339; IPCC assessments
338–339; value-ladenness of climate assessment
reports 339–345; values and assessment reports
on 337–345
- climate change frames: science and values matter
for 425–432; and values 428–430; *see also*
framing
- climate science 177–178, 181–182, 241, 306,
309, 311, 338, 342–344, 394; consequences
483–486; dealing with values in 483–486;
and inductive risk 479–480; integrated
assessment models 481–483; non-epistemic
values in global climate modeling 480–481;
philosophical approaches to values in 478–486;
philosophical responses 483–486; strategies
483–486
- Clostridium botulinum* 377

- Coates, Matthew 318–319
 Coca-Cola 295
Coccoidea 420–421
 Code, Lorraine 29
 Coen, Deborah 391
 coercive policies 367–368
 cognitive diversity 163
 cognitive mobilization 232
 cognitive values 28, 30, 87; *see also* epistemic values
 collaboration: intercultural 413–422; scientific 446–447
 Collins, Frances 378
 communication (*see science communication*)
 communicative action 136
 communism 298, 519
 community-based participatory research (CBPR) 202–203; *see also* citizen science, participant-led science, public participation model, stakeholder involvement
 COMPAS controversy 264, 266
 compliance 287, 289
 conceptual argument 54–55; *see also* framing, mixed claims
 consensus 134–136; failure of 136–139
 consequences: of being measured 393–394; of error 392–393; of metrological standards 391–392
 consequentialist demarcation strategies 305; *see also* demarcation strategies
 conservation biology 489, 491, 496–497; *see also* biology, wildlife conservation research
 conservation outreach 495–496
 consistency 284–285
 constraint condition 203
 constraints, EPA 123
 content of measurands 395–398
 contingency argument 55–56; *see also* unforced choices
 Contract Research Organizations (CROs) 298
 coordinative demarcation strategies 305; *see also* demarcation strategies
 corporations 7, 132, 194, 213, 257, 294
 cost-effectiveness 286
 Cottam, Hilary 508
 Council of Tobacco Research (CTR) 306
 coupled epistemic/ethical issues 13
 COVID-19 pandemic 2, 11, 378–379; *see also* SARS coronavirus, severe acute respiratory syndrome
 Crazy Bull, Cheryl 351
 criminal justice algorithms 268
 Critical Contextual Empiricism (CCE) 9, 141; early critiques of 142–144; expansion and application of 147–148; overview 142; recent critique of 144–147; *see also* transformative criticism
 Crouch, Martha 257
 cultural anchor 236
 Cures Act 211, 213
 Curry, Tommy J. 270–271
 Daston, Lorraine 42, 43
 Daszak, Peter 378
 data: characterizing 200–203; completeness 164; diversity 162, 164–167, 170; generation 267–268; and research purpose 205–206
 da Vinci, Leonardo 86
 debates on science and values 327–328
 decision-making 268–269
 decolonial scholarship 415
 decolonization in politics of knowledge 414–416
 deferred decision response 65–66
 deficit model of science communication 436–437
 Deloria, Ella 351
 Deloria, Vine, Jr. 351
 demarcation strategies 305; *see also* new demarcation problem (NDP)
 de Melo-Martín, Inmaculada 69–70, 116, 147
 democracy 134–136; is rooted in emancipative values 231–232; political shape of 233; and science 9–10, 220–228; value judgments 232–235; and values 9–10
 democratic defense 66–67
 democratic ideals 152
 democratic pluralism 235–239
 democratic values 67, 181, 193, 204, 311
 Department of Health and Human Services (DHHS) 374–377
 design/conduct of studies 296–298
 Desmond, Hugh 448–449
 detached objectivity 25
 Dewey, John 29, 221
 diagrams 340–342
 Diamond Open Access 157, 158n5
 Dichlorodiphenyltrichloroethane (DDT) 306
 Dierickx, Kris 448–449
 Dirac, Paul 86
 direct and indirect roles for values in science 315–316
 disability-adjusted life years (DALY) 397
 disinformation 287; *see also* misinformation
 dissent 318–319
 diversity 414–416; in biomedical databases 164–165; cognitive 163; data 162, 164–167, 170; epistemic 163, 165, 170, 414–416, 421; within extended scientific community 194–195; geographic 151–158; ineffective epistemic 156; limits in science 162–170; in microbiome research 165–166; revision of notions of 169; social 163, 169, 414; social value 163; value of 163–166
 diversity, equity, and inclusion (DEI) 282, 285
 Dokic, Jérôme 89
 Douglas, Heather 2, 3–6, 25, 52–53, 58–59, 64, 65, 66, 96, 111, 132–133, 196n8, 403, 431, 436
 Dror, Lidal 38
 dual use biomedical research 374–383; Biden Administration policy on 379–380; COVID-19 pandemic 378–379; DHHS policy 377; dual use

- research 374, 380–382; H5N1 avian flu research 376–377; potential pandemic pathogens 379–380; social values 380–382; Trump administration on gain-of-function research 380; *see also* [biomedical research](#)
- dual use research (DUR) 374; and social values 380–382
- dual use research of concern (DURC) 374, 377, 379–383
- Du Bois, W.E.B. 66–67, 132
- Duhem, Pierre 86
- economics: classical 501–502, 503–505; neoclassical 501–502; values in 501–509; values of Protestant work ethic 502; *see also* [heterodox economics](#), [neoclassical economics](#)
- Edenhofer, Ottmar 65
- education and advising 452; *see also* [science education](#)
- Edwards, Robert 234
- efficiency 284
- Elabbar, Ahmad 74, 225, 485–486
- Elgin, Catherine 88
- eliminativism 402–403
- eliminativists 401–410
- Elliott, Kevin C. 23, 56, 58, 65, 77–78, 110, 115, 131, 176, 179–180, 200, 206, 246, 286, 312n2, 327
- embryo proper 234
- empirically informed philosophical approaches 326–333
- “Empiricism and the Philosophy of Mind” (Sellars) 74
- energy transitions 426–428
- environmental ethics 351; *see also* [biomedical ethics](#), [research ethics](#)
- environmental justice/injustice 56, 355, 430
- environmental science assessments: current interventions in 355–356; exclusion of Indigenous Peoples from 353–355; Indigenous peoples’ leadership in 348–359; *see also* [assessment](#)
- epistemic aims of research 446
- epistemic authority 339
- epistemic curation 343–345
- epistemic diversity 163, 165, 170, 414–416, 421
- epistemic exploitation 153, 155, 156, 163
- epistemic goals 123
- epistemic inequality 194
- epistemic intimidation 305, 306–307, 312n3; contemporary antecedents 305–306; effects of 309–311; and illegitimate value-influences in science 304–312
- epistemic justice/injustice 13, 38, 153, 306, 321, 394
- epistemic justification of research ethical norms 445–448
- epistemic marginalization 153, 156
- epistemic oppression 152, 153–154
- epistemic powerlessness 153–154, 155
- epistemic projection approach (EPA) 70, 120–121; defined 123; foundations of 121–122; in practice 125–128; problem-selection stage of 122–124, strengths and limitations of 128–129; values in problem selection 122–125
- epistemic research problem 127
- epistemic responsibility 447
- epistemic risks 7; *see also* [inductive risk](#), [ontic risk](#), [representational risk](#)
- epistemic toolbox 136
- epistemic trade-off 166–167
- epistemic values 2–6, 25, 28, 43, 45–46, 48–49, 52–53, 59, 63, 115, 136–137, 142, 166–167, 304, 444, 447; plurality of 450; *see also* [cognitive values](#)
- epistemic virtues: and problem of inductive risk 307–308; and vices 312n2
- epistemology: civic 235–236; moralized 42
- Epstein, Steven 164, 458
- equal participation and leadership 357–358
- error and silence 339–340
- ethical review 452–453
- ethical values 2–4, 25, 29, 52–53, 58, 64, 67, 133, 209, 252, 319, 377, 403, 430–431; *see also* [moral values](#), [non-epistemic values](#), [public values](#), [religious value](#), [social values](#)
- ethics 48–49, 269–274; approval 334n5; biomedical 13; environmental 351; insurrectionist 270–271; work 503–505
- ethnicity 164, 169–170, 401, 468–470, 474–475, 514
- EU AI Act 517–518
- European Citizen Science Association 440
- European Code of Conduct for Research Integrity* 449
- evidence-based policymaking (EBPM) 370
- explanandum 167
- explanans 167
- extractivism 154–156; and completeness 168
- Fabian, Mark 204
- fairness 10, 224, 263–270, 284, 285, 446, 515, 518–520
- fairness metrics 264–265
- faithfulness to facts 42–44
- Fanon, Frantz 44
- The Fate of Knowledge* (Longino) 146
- Fauci, Anthony 378
- Fazelpour, Sina 321, 322
- feminism: alternatives to VFI 29–30; conceptions of objectivity 29–30; in feminist standpoint theory 39; science question(s) in 23–25; *see also* [gender](#), [sex](#), [sex difference research](#), [sexism](#)
- feminist critics 459–461
- feminist standpoint empiricism 40
- feminist standpoint theory 34–35, 38–40
- Fernández Pinto, Manuela 298
- Feyerabend, Paul 236

- Floyd, George 266
 Fouchier, Ronald 376
 framing: new demarcation problem (NDP) 110;
 problem 512–513; self-sacrifice 429; Tragedy of
 the Commons ToC 425–426; value-laden 489–
 498; *see also* climate change frames, conceptual
 argument, self-sacrifice framing, ToC framing,
 value-laden framing
 Franco, Paul 342
 Franklin, Allan 88
 Freedom Convoy 363, 369, 371n1
 French, Steven 84
 Frisch, Mathias 65, 482
 Fryer, Daniel 270, 272–274
 functional beauty 90
 functionalist demarcation strategies 305; *see also*
 demarcation strategies
 fundamental vs. applied science in dynamic world
 496–497
 funding 154–156, 281–285; *see also* industry-funded
 research, philanthrocapitalism, philanthropic
 science, public science funding, research agendas
 funding effect 296; *see also* industry-funded research
 Future Crops Training Centre 331
- gain-of-function research 380
 Galilei, Galileo 85
 Galison, Peter 42, 43
 gap argument 53
 gender 24–26, 36, 40, 163, 190, 514; -based
 discrimination 37; bias 91; equality 77; identity
 258; inequality 285; norms 27; stereotypes 457,
 459–461, 464n3; studies 304, 306; *see also*
 feminism, sex
 genome-wide association studies (GWAS) 467;
 see also biomedical research, genomics, human
 genomics
 genomics 11–12, 162, 164–166, 168; research
 326–333; values in human 467–475; *see also*
 biomedical research, genome-wide association
 studies (GWAS), human genomics
 geographic diversity 151–158
 ghostmanagement 299
 ghostwriting 299
 global climate modeling 480–481; *see also* modeling
 global environmental assessments (GEAs); *see also*
 assessment 337
 global science: structural inequalities of 154–157;
 values in 151–158
 Global South 152–153, 156, 158, 158n1, 166, 190,
 225, 288, 294, 415
 glomerular filtration rate (GFR) 401–402
 goal-relativity of response-requiring criticism (GRC)
 principle 144–145
 Goldman, Alan 146
Goniopsis cruentata 419
 Government agencies
- gradualism 271–274
 Grandjean, Philippe 296
 graphs 340–342
 ground truth 267–268
- H3Africa 468, 471, 473
 H5N1 avian flu research 376–377
 H5N1 avian influenza virus 376
 Habermas, Jürgen 136, 225, 235–239
 Hacking, Ian 47
 Hansson, Sven Ove 113
 Haraway, Donna 29
 Hardie, Jeremy 46
 Hardin, Garrett 425
 Harding, Sandra 29, 47–48
 Harris, Leonard 270–271
 Harvard, Stephanie 7
 Haslanger, Sally 47
 Havstad, Joyce 53–54, 65
 Hawkesworth, Mary 47
 hedging response 65–66
 Heisenberg, Werner 86–87
 Hellman, Deborah 264–266
 Hempel, Carl G. 95–96, 308, 410n1
 Henry Roe Cloud 351, 354
 Henschen, Tobias 66
 heterodox economics 505–507; *see also* economics
 heterodox economists 506–507; *see also* economics
 Hicks, Dan 109
 Hilligardt, Hannah 117
 Hobma, Marcel 117
 Holland, Mark 363
 Holman, Bennett 6, 8–9, 58, 64, 109, 111, 305, 311,
 317–319, 322, 520n8
 Hooke, Robert 86
 Hossenfelder, Sabine 87
 Hrdy, Sarah 27
 human communities 255–256
 Human Development Index (HDI) 507
 human genomics: *All of Us* cohort study 473–475;
 human genomics be value-free 468–471; slippage
 469–470; underdetermination 470–471; values
 and alternative hypotheses/conceptualizations
 472; values guiding choice of research goals 473;
 values helping design research 471–472; values
 in 467–475; values managing inductive risk 472–
 473; *see also* biomedical research, genomics
 human microbiome research 162–170; *see also*
 biomedical research
 human subjects 254–255; *see also* research ethics
 human welfare 507–509; capabilities approach
 to 507–509; progressive work ethic
 507–509
- ideal theory 67–68
 ideology 244–247
 ignorance 38, 147, 282, 296; *see also* agnotology

- illegitimate value-influences in science 304–312; *see also* [new demarcation problem](#)
- incrementalism 264–265; fairness metrics 264–265; modular reform 264–265
- incrementalist paradigm 263, 265
- Indigenous Elders 353
- Indigenous groups 168, 255–257, 417, 489
- Indigenous Knowledges 350, 352
- Indigenous-led assessments 358–359; *see also* [assessment](#)
- Indigenous Peoples: equal participation and leadership 357–358; Indigenous-led assessments 358–359; leadership in environmental science assessments 348–359
- Indigenous Peoples' Customary Governance for Sustainable Management of Natural Resources and Protection of Biodiversity* 358
- inductive risk 7, 95–97, 120, 307–308, 406–409, 450–451; argument from 403–404; and climate science 479–480; downstream regress of 65; upstream regress of 65; values introduced to manage 472–473; *see also* [epistemic risk](#), [ontic risk](#), [representational risk](#)
- Industrial Revolution 232, 503
- industrial selection 316–318
- Industrial Transformation Training Centre (ITTC) 329
- industrial withholding 321–322
- industry 7, 11; bias 297; defined 334n2; editorial 157; groups or organizations 296; private 155, 287; sponsorship 143
- industry-funded research: agenda setting 295–296; communicating scientific findings 298–299; design and conduct of study 296–298; potential solutions 299–300; use of scientific categories 296–298; values and 294–300; *see also* [funding](#), [industry-funded science](#), [tobacco strategy](#)
- industry-funded science 195, 294–300; *see also* [funding](#), [industry-funded research](#), [tobacco strategy](#)
- ineffective epistemic diversity 156
- inequality: epistemic 194; gender 285; structural 154–157
- injustice: environmental 56, 355, 430; epistemic 13, 38, 153, 306, 321, 394
- inquiry (*see* [scientific inquiry](#))
- Institute of Electrical and Electronics Engineers (IEEE) 511
- institutional deployment 268–269
- institutionalization and political constitution 338–339
- institutional policies: education and advising 452; ethical review 452–453; increasing awareness of values in science 451–453; public commitment 451–452; sanctions for norm violations 452
- institutions: and science 10–11; and values 10–11; *see also* [organizations](#), [science funding institutions](#)
- insurrectionist ethics 270–271
- insurrectionist ethics vs. anti-ethics 270–271
- integrated assessment models (IAMs) 481–483
- intellectual beauty 89, 90
- Intemann, Kristen 2, 56, 69–70, 483
- Inter-American Indian Conference 354
- Inter-American Indian Institute 354
- interdisciplinarity (*see* [transdisciplinarity](#))
- intercultural collaboration 413–422
- Intergovernmental Panel on Climate Change (IPCC) 11, 309, 337, 348; institutionalization and political constitution 338–339; primer on 338–339; SROCC 341
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) 348
- International Coral Reef Initiative (ICRI) 356
- International Indigenous Forum on Biodiversity and Ecosystem Services (IIFBES) 356
- International Labour Organization 349
- integrity (*see* [scientific integrity](#))
- Intertribal Council on Utility Policy 356
- Irzik, Gurol 182
- Ivanova, Milena 84, 89, 90–91
- Ivins, Bruce 375
- Jasanoff, Shelia 232–235
- Jefferson, Thomas 223
- Jeffrey, Richard C. 65
- Jim Crow 264
- Joel, Daphna 463
- Jofre-Bonet, Mireia 298
- John, Stephen 113, 227, 228n1, 242
- Jordan-Young, Rebecca 461
- journalism (*see* [media](#))
- justice 10, 52; -based analyses 394; environmental 56, 355, 430; epistemic 13, 38, 153, 306, 321, 394; social (*see* [social justice](#)); *see also* [structural inequalities of global science](#)
- Justice as Fairness* (Rawls) 266
- Kari-Oca Declaration 355
- Kelleher, J. Paul 482
- Kitcher, Philip 46
- knowledge: extraction 415; mining 415; moral 135; systems 12; *see also* [scientific knowledge](#)
- Kogelmann, Brian 135
- Kojevnikov, Alexei 241, 247
- Kopec, Matthew 429
- Korf, Rebecca 77–78, 312n2, 317, 319
- Koskinen, Inkeri 110, 115
- Kotkin, Stephen 245
- Kowarsch, Martin 65
- Kuhn, Thomas 2, 28, 39, 43, 45, 78
- Kukla, Quill 104, 328
- Kurtulmus, Faik 182
- Kyoto Protocol 428–429

- Lafont, Cristina 231–232
 large language models (LLMs) 511
 Lari, Teemu 144–145
 Latour, Bruno 226
 Laudan, Larry 3
 Lawson, Cornelia 298
 LEAP (Lifetime Exposures and Asthma outcomes Projection) 98
 LEAP-SIMPLE model 98–102, 99–100, 104n2
 learning from partial overlaps 418–421
 legal resources 312
 legitimacy (*see* [scientific legitimacy](#))
 Leonelli, Sabina 158
 Levi, Isaac 43
 Lewens, Tim 55
 liberal state 235–239
 Lira, Ignacio 392
 list-based solutions 115–116
 lithium batteries 430
 Livingston, Jasmine E. 344
 Lloyd, Elizabeth 45, 46
 local diversities 168–170
 local knowledge 223, 352, 356, 415–418
 Lombrozo, Tania 88
 Longino, Helen 2, 29, 35, 37, 45, 47–48, 141–143, 328, 457
Lost in Math: How Beauty Leads Physics Astray (Hossenfelder) 87
 Ludwig, David 55, 352
 Lusk, Greg 56, 67
 Lysenko, Trofim 242–243

 machine learning 10, 166, 263, 264, 265, 267, 269, 513
 MacKinnon, Catharine A. 44
 Macron, Emmanuel 369
 Madison, James 223
 Magnus, P. D. 79–81
 Malthus, Thomas 504
 mandatory vaccination 371n6
 Mann, Michael 305
 Māori territorial self-governance 351
 Martin, Emily 27
 Marx, Karl 243, 505, 509
 Mathews, John Joseph 351
 McAllister, James W. 84, 87
 McClimans, Leah 393–394
 McDowell, John 75
 McGowan, Mark 369
 McKaughan, Daniel J. 56, 58, 179–180, 245–246
 McMullin, Ernan 3, 75
 measurement: background assumptions behind 389–390; consequences of being measured 393–394; consequences of error 392–393; consequences of measurement 390–395; consequences of metrological standards 391–392; content of measurands 395–398; justice-based analyses 394; values and 389–398; values and validity 394–395
 media: mainstream 378; social 148, 306, 312, 393; *see also* [science communication](#)
 medical freedom 364, 368, 370
 Menon, Tarun 68–69, 133
 Menzies, Charles 259
 Merton, Robert 220, 298
 Messick, Samuel 394
 Michaels, David 299
 Mills, Charles 225, 263, 265–267, 275
 Mishel, Larry 506
 misinformation 1–2, 7–8, 147, 225, 227, 320, 364, 369; *see also* [disinformation](#)
 Mitchell, Sandra 65
 mixed claims 54, 179; *see also* [conceptual argument](#)
 modeling 7, 11–12, 66, 97–98, 102–103, 104n1, 253, 350–351; *see also* [global climate modeling](#), [network models](#)
 modular reform 264–265
 molecular biology 329, 331; *see also* [biology](#), [biomedical research](#)
 Montano, Uliyanov 87
 Montuschi, Eleonora 46
 moral imagination 430–432
 moralized epistemology 42
 moral justification of research ethical norms 445–448
 moral knowledge 135
 moral values 448–451; *see also* [ethical values](#), [non-epistemic values](#), [public values](#), [religious value](#), [social values](#)
 Mouffe, Chantal 227
 Moynihan, Daniel 112
 Mukerji, Chandra 281–282
 Murphy, Alice 84, 85

 Nagel, Thomas 43
 narcissists 154
 narratives 497–498
 National Academies of Science, Engineering and Medicine (NASEM) 468
 National Academies of Sciences 471
 National Aeronautics and Space Administration (NASA) 356
 National Institutes of Health (NIH) 76–77, 309, 374
 National Research Council (NRC) 375
 National Science Advisory Board for Biosecurity (NSABB) 375
 Native American Graves Protection and Repatriation Act 256
 natural science 378–379; *see also* [science](#)
 Navin, Mark C. 367–368
The Neglect of Experiment (Franklin) 88
 neoclassical economics 501–502, 505–507; *see also* [economics](#)
 neoclassical economists 506; *see also* [economics](#)

- network models: dissent 318–319; industrial selection 316–318; industrial withholding 321–322; open science movement 319–320; out-group mistrust 320–321; public communication 320; values and 316–322; *see also* modeling
- networks and values 315–322
- new demarcation problem (NDP) 6, 8–9, 109–117; axiological approaches 114–115; categorization 110–111; framing 110; functionalist responses 113–114; reframing 116–117; *see also* demarcation strategies, illegitimate value-influences in science
- New Zealand Climate Change Programme 351
- Nguyen, C. Thi 177–178
- NIH Revitalization Act 458
- Nineteen Eighty-Four* (Orwell) 245
- NKF-ASN Task Force 406
- non-epistemic trade-offs 167–168
- non-epistemic values 2–5, 8–9, 12, 23, 25–30, 34, 42–43, 45, 49, 55–57, 63–64, 66–69, 74–77, 79–82, 110, 115, 133, 136–137, 166, 175–183, 209, 263, 403–405, 419, 436, 438, 448–451, 471–475, 480–481; *see also* ethical values, moral values, public values, religious value, social values
- non-ideal theory: of algorithmic governance 275–276; of scientific legitimacy 274–276
- non-medical exemptions (NMEs) 367–369
- no-passing-the-buck principle 65
- Nordhaus, William 483
- Nussbaum, Martha 508
- objectivity: detached 25; as faithfulness to facts 42–44; feminist conceptions of 29–30; intertwined criticisms 44–46; as value-freedom 42–44; and value-free ideal 25–27; value-laden 47–48
- Okruhlik, Kathleen 58, 70
- O'Neill, Maggie 270, 271, 274
- ontic risk 7; *see also* epistemic risk, inductive risk, representational risk
- ontology of values (*see* values, ontology of)
- Open Insulin Foundation (Open Insulin project) 438
- openness 144–145; *see also* open science, transparency
- openness of goals to criticism (OGC) principle 144–145
- open science 9, 176, 298, 319–320, 415, 439–440, 452; *see also* openness, transparency
- organizations: and science 10–11; and values 10–11; *see also* institutions
- Orwell, George 245
- out-group mistrust 320–321
- Paris Agreement 429
- Parker, Wendy 70, 484
- Parsons, Glenn G. 90
- partial overlaps: learning from 418–421; in value commitments 416–418; and value-ladenness of science 413–414
- participant-led science 438–440; *see also* citizen science
- partnerships, in agricultural genomics research 326–333; *see also* research partnerships, university-industry research partnerships
- Pasang Dolma Sherpa 353
- paternalism 414–416
- patient-reported outcome measures (PROMs) 393
- Patt, Anthony 427
- Paul, Rand 378
- Paul K. Feyerabend Foundation (PKFF) 288
- perceptual beauty 89; *see also* aesthetic values
- personal probabilities 393
- Peter, Fabienne 248
- Pfizer 285
- philanthrocapitalism 288; *see also* funding
- philanthropic science 287–289; accountability 288–289; stakeholder involvement 287–288; sustainability and philanthrocapitalism 288; trust and compliance 289; *see also* funding
- philosophical approaches: in climate science 478–486; to values 478–486
- physics 12–13, 49, 87, 89, 91, 110, 142, 246–247, 253, 295
- Piketty, Thomas 508
- Planck, Max 389
- Poincaré, Henri 86
- Polanyi, Michael 223, 235
- policy making and policy makers 3, 7, 11, 13, 54, 391, 485; *see also* vaccines, policy of
- political constitution 338–339
- political culture 233–234
- political legitimacy 247–248
- Political Liberalism* (Rawls) 266
- political philosophy of science 6, 225–227
- The Politics of Evidence-Based Policymaking* (Cairney) 370
- politics of knowledge 414–416
- Porter, Theodore 398n1
- power: of name 493–494; of value-laden framing 489–498
- pragmatic experimentalism 273
- pre-embryo 234
- preferences, EPA 123
- Prinz, Michael 429
- private science funding (*see* industry-funded research)
- problem selection (*see* epistemic projection approach (EPA), problem-selection stage of)
- Proceedings of the National Academies of Sciences* (PNAS) 375–376
- progressive work ethic 507–509
- Protestant work ethic 502
- Proxmire, William 283
- pseudorationalism 82n4
- psychology 24, 78, 110, 132, 190, 272, 452, 461

- public commitment 451–452
 public communication 320; *see also* [science communication](#)
 public good 103, 111, 191–192, 195, 252, 257, 331, 334n8, 426, 429, 431
 The Public Justification Principle 134
 public participation model 437–438; *see also* [community-based participatory research](#)
 publics 111, 177, 181, 183, 215, 234, 244, 316, 328, 332, 334n4, 364–369, 437–440, 441n5
 public science funding 281–285; accountability 282; autonomy 282–283; consistency 284–285; diversity, equity, and inclusion (DEI) 285; efficiency 284; fairness 284; reliability 284; transparency 283; trust 283; *see also* [funding](#)
 public values 199–206; *see also* [ethical values](#), [moral values](#), [non-epistemic values](#), [religious value](#), [social values](#)
 quality-adjusted life years (QALY) 397
 Quine, Willard V. O. 28
 race 167–168; adjustment 404–409; correction 401–403
 Racial Algorithm 263–276; compounding effects 269; data generation 267–268; decision-making 268–269; decisions 269; defining 267; distributions 269; institutional deployment 268–269; overview 263
 Racial Contract 263, 266–269
 racism (*see* [scientific racism](#))
 randomizer reply 78–80, 82n4
 Rawls, John 225
 Reducing Emissions from Deforestation and Forest Degradation (REDD) 349
 regulation 10, 176, 179, 329, 518
 Reichenbach, Hans 25
 reliability 1, 3, 57, 132, 134, 282, 284, 294, 296, 305, 338, 392, 446, 518
 religion: in democratic society 231–239; overview 231
 religious value 10; *see also* [ethical values](#), [non-epistemic values](#), [moral values](#), [public values](#), [social values](#)
 representational risk 7, 97–104; *see also* [epistemic risk](#), [inductive risk](#), [ontic risk](#)
 representativeness condition 203
 reproduction: behavior beyond 461–462; of hierarchy 268
 research agendas 154–156; *see also* [funding](#)
 research ethics 8, 13, 46, 69, 204, 444–453; *see also* [biomedical ethics](#), [environmental ethics](#), [human subjects](#), [selective reporting](#)
 research evaluation 156–157
 research partnerships: in Australian landscape 328–329; science and values debates 332–333; university-industry partnership 330–332; *see also* [partnerships](#), in [agricultural genomics research](#), [university-industry research partnerships](#)
 research purpose: and data 205–206; understanding 203–205
 research stage, of EPA 124–125
 Resnik, David B. 110, 115
 risk: epistemic 7; inductive (*see* [inductive risk](#)); ontic 7; representational 7, 97–104
 Ritson, Sophie 89
 Robles-Piñeros, Jairo 420
 Rolin, Kristina 110, 115
 Rooney, Phyllis 57
 Rorty, Richard 75
 Roth, Aaron 264, 266
 Rudner, Richard 53, 65, 66, 95, 403
 Rueger, Alexander 90
 Rummukainen, Markku 344
 Rutherford, Ernest 86
 Said, Edward 44
Salmonella 375, 381
 sanctions for norm violations 452
 SARS coronavirus (SARS-CoV) 377; *see also* [COVID-19 pandemic](#)
 Scheman, Naomi 30
 Schroeder, Drew 204, 224
 science: aesthetics of 89–91; awareness of values in 451–453; de-idealizing political philosophy of 225–227; and democracy 9–10, 220–228, 231–239; diversity limits in 162–170; vs. freedom 369–370; geographic diversity on 152–153; human subjects 254–256; industry-funded 195, 294–300; institutional policies 451–453; and institutions 10–11; managing values in 8–9; non-epistemic values in 178–181; and organizations 10–11; overview 1–2; partial overlaps 413–414; participant-led 438–440; philosophy of 223–225; and race correction 401–410; role in liberal democracies 222; and social justice 252–259; specific, values in 12–13; theoretical background on 8; transparency in 175–183; value-laden 74–81, 413–414; value question(s) in 27–29; values in 315–316, 401–410, 430–432; *see also* [natural science](#)
Science and Moral Imagination (Brown) 431
 science and technology studies 8, 13, 226, 328, 338, 428
Science as Social Knowledge (Longino) 142
 science communication 493–496; deficit model of science communication 436–437; participant-led science 438–440; rise of public participation model 437–438; values in 435–440; Western 435; *see also* [public communication](#)
 science education 117, 351; *see also* [education and advising](#)
 science for sale 285–287; *see also* [industry-funded research](#)

- science funding institutions: philanthropic science 287–289; public science funding 281–285; science for sale 285–287; values of 281–289; *see also* institutions
- scientific assessments 7, 11, 74, 339, 341, 350–351, 354–356, 358; *see also* assessment
- scientific collaboration 446–447
- scientific inquiry 448–451
- scientific integrity 286; *see also* research ethics
- scientific/intellectual movements (SIMs) 39
- scientific knowledge 135, 222; as public resource 189–195; science and public need 191–195
- scientific legitimacy 274–276
- scientific racism 469
- selective reporting 297–300; *see also* research ethics
- self-sacrifice framing 429; *see also* framing
- Sellars, Wilfrid 74
- Sen, Amartya 508
- September 11, 2001 attacks 375
- servility 154–155
- severe acute respiratory syndrome (SARS) 375, 377–379; *see also* COVID-19 pandemic
- sex 146, 164, 402, 462; chromosomes 24; differences 24–25; essentialism 12, 457–464; superiority of 460; *see also* feminism, gender, “sex as a biological variable” (SABV)
- “sex as a biological variable” (SABV) 457, 464n2; *see also* biology, biomedical research
- sex difference research and feminist critics 459–461; *see also* feminism
- sexism 148, 253, 258, 304, 308, 457, 460; *see also* feminism
- Sheykh-Rezaee, Hossein 68
- Silent Spring* (Carson) 306
- situated knowledge thesis 35–37
- slippage 469–470
- Smith, Adam 503–505, 508–509
- Social Contract for Science (SCS) 436
- social cost of carbon (SCC) 481
- social diversity 163, 169, 414; *see also* diversity
- social integration 236
- social justice: human communities 255–256; human subjects 254–255; overview 252–253; research agenda 256–258; research ethical norms 445–448; and science 252–259
- social media 148, 306, 312, 393; *see also* science communication, media
- social norms 364, 367
- social values 448–451; in core activities of scientific inquiry 448–451; diversity 163; and dual use research 380–382; *see also* ethical values, non-epistemic values, moral values, public values, religious value
- sociology 85, 201
- Soleiman, Matthew 46
- space: of causes 74–75; of reasons 74–75
- STACC Report 359
- stakeholder involvement 287–288; *see also* community-based participatory research
- Stamenkovic, Philippe 113
- Standards of Educational and Psychological Testing 394
- standpoint theory: achievement thesis 38–39; ideals 152; overview 34–35; situated knowledge thesis 35–37; thesis of epistemic advantage 37–38
- Stanford Encyclopedia of Philosophy* (SEP) (Hellman) 264
- Status of Tribes and Climate Change Report *see* STACC Report
- Steel, Daniel 65, 321, 322
- Stegenga, Jacob 53, 68–69, 133
- Stein, Zachary 394
- Steptoe, Patrick 234
- stereotypes 167–168
- Stich, Stephen G. W. 135
- strategy/ies 137–139
- structural inequalities of global science 154–157; academic migration 156; academic publishing 156–157; extractivism 154–156; funding 154–156; research agendas 154–156; research evaluation 156–157
- structural inequality 154–157
- structures or structural analyses 266–267
- subordinate integration 154
- Summary for Policymakers (SPM) 339
- Sung, Doohyun 113
- support networks 311, 312
- sustainability 287, 288, 519; discourses 417–418; ecological 417; environmental 4, 52, 332; sustainability-related ideas 417
- systemic approach 9
- systemic bias 87–88
- systemic demarcation strategies 305; *see also* demarcation strategies
- taxonomy-guided assessment 80–81; *see also* assessment
- technosciences 328
- A Theory of Justice* (Rawls) 266
- thesis of epistemic advantage 37–38
- Thresher, Ann C. 46
- timeliness 285, 286
- timidity 154, 157
- tobacco strategy 306; *see also* industry-funded research
- ToC framing 425–426; *see also* framing
- Torsell, Christian 319
- tragedy of the commons (ToC) 425, 426–428
- training 312
- transdisciplinarity 414–415
- transformative criticism 2, 29; *see also* Critical Contextual Empiricism (CCE)
- transparency 181–183, 183n2, 283; aims 176–178; costs 176–178; methods 176–178; in science 175–183; *see also* openness, open science

- transpolitical authority 114
- Trudeau, Justin 363
- Truman, Harry S. 414
- Trump, Donald 378, 380
- trust 181–183, 283, 287, 289
- Tunks, Andrea 351
- Turner, Erick H. 84
- uncertainty 4, 35, 53, 65–66, 96, 102–103, 121–122, 148, 215, 294, 306, 327, 338–340, 392
- underdetermination 405–406, 470–471
- unforced choices 199
- United Nations Environment Program (UNEP) 350
- United Nations Framework Convention on Climate Change (UNFCCC) 350, 353, 428
- “university-industry research partnerships” 333n1
- US 2009 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPIcr) 402
- US Defense Advanced Research Projects Agency (DARPA) 379
- US Fish and Wildlife Service (USFWS) 111–112
- US Food and Drug Administration (FDA) 458; Accelerated Approval program 213; challenges incorporating stakeholders’ values 212–215; stakeholders’ values 210–212
- U.S. Global Change Research Program (USGCRP) 357
- U.S. National Climate Assessment 356, 357
- US National Institute of Standards and Technology (NIST) 517–519
- vaccine mandates 366–367; as coercive policies 367–368; to establish social norms 367
- vaccine policy: controversies in 363–370; science vs. freedom 369–370; vaccine mandates 366–367; vaccine mandates as coercive policies 367–368; vaccine mandates to establish social norms 367; vaccine policies 365–368; vaccines and values 364–365; *see also* policy making and policy makers
- validity 394–395
- Vallier, Kevin 134
- value commitments 416–418
- value-freedom 42–44, 134–136
- Value-Free Ideal (VFI) 2–6, 9, 63–64, 120, 436; as action guidance 132–134; alternatives to 29–30; arguments against 52–59; arguments for 64–69; feminist challenges to 23–30; and objectivity 25–27; partial rapprochements with 69–71; and value-management questions 64
- value judgments 200, 232–235
- value-laden background assumptions 449
- value-laden framing 489–498; advancing the narrative 497–498; conservation context 489–491; conservation policy and impact 496–497; power of 489–498; during research and analysis 491–492; in science communication 493–496; *see also* framing
- value-ladenness: causal interpretation of 76; of climate assessment reports 339–345; motivational interpretations of 76; objectual interpretation of 77; ontology of values 77–78; rational interpretations of 76; senses of 75–77
- value-laden objectivity 47–48
- value-management: intermediate approach to 120–129; questions 64
- value neutrality 148
- value overlaps: in intercultural collaboration 413–422; in transdisciplinary 413–422
- values: aesthetic 49, 84–91; alternative hypotheses and conceptualizations 472; and assessment reports on climate change 337–345; and climate change frames 428–430; for climate change frames 425–432; and democracy 9–10; and dual use biomedical research 374–383; in economics 501–509; geographic diversity on 152–153; in global science 151–158; guiding choice of research goals 473; helping design research 471–472; heterodox economic perspectives on 505–507; in human genomics 467–475; and industry-funded research 294–300; and institutions 10–11; in integrated assessment models 481–483; managing 5–6; managing, in science 8–9; managing inductive risk 472–473; and measurement 389–398; nature of 3–5; neoclassical economic perspectives on 505–507; and network models 316–322; and networks 315–322; ontology of 77–78; and organizations 10–11; overview 1–2; of Protestant work ethic 502; and race correction 409–410; in research ethics education 444–453; in science 315–316; in science and moral imagination 430–432; in science communication models 435–440; of science funding institutions 281–289; in science in nondemocratic contexts 241–249; in scientific activities 11–12; in specific sciences 12–13; theoretical background on 8; and vaccines 364–365; vaccines and 364–365; and validity 394–395; *see also* ethical values, moral values, non-epistemic values, public values, religious value, social values
- values conflicts 205
- van den Wijngaard, Marianne 463–464
- vanity 154, 155
- Vavilov, Nikolai 243, 247
- Vavilov, Sergei 247
- Verhaeghe, Pieter-Paul 310
- visual representation 340–342
- vulnerable populations argument 272
- Wagner, Wendy 116
- Walker, David 270
- Ward, Zina B. 7, 77–78
- well-being 4, 28, 46, 54–55, 74, 179–180, 393–394, 396–397, 479–480, 501–502

Index

- well-ordered science [224–226](#), [228](#)
- Welzel, Christian [232](#), [235–236](#)
- wildlife conservation research [489–498](#); *see also*
 - [conservation biology](#)
- Wilholt, Torsten [6](#), [8–9](#), [64](#), [67](#), [109](#), [111](#), [305](#),
[311](#), [520n8](#)
- Williams, Bernard [43](#), [248](#)
- Winsberg, Eric [7](#), [65](#)
- work: ethics [503–505](#); heterodox economic
 - perspectives [505–507](#); neoclassical economic
 - perspectives [505–507](#)
- World Health Organization (WHO) [288](#), [364](#), [371n3](#)
- World Meteorological Organization (WMO) [350](#)
- worldview neutrality [237](#), [239](#)
- World War II [414](#), [436](#)
- Wuhan Institute of Virology (WIV) [377–379](#), [381–382](#)
- Wylie, Alison [34–36](#), [84](#)
- Yoshihiro Kawoaka [376](#)
- Young, Iris Marion [226](#)
- Zhengli, Shi [377](#)
- Zollman, Kevin J.S. [317–319](#)
- Zollman effect [322](#)