

History, Philosophy and Theory of the Life Sciences

João L. Cordovil
Gil Santos
Davide Vecchi *Editors*

New Mechanism

Explanation, Emergence and Reduction

fct Fundação
para a Ciência
e a Tecnologia

OPEN ACCESS

 Springer

History, Philosophy and Theory of the Life Sciences

Volume 35

Series Editors

Philippe Huneman, (CNRS/Université Paris I Panthéon-Sorbonne),
Institut d'Histoire et de Philosophie des Sciences et des Techniques,
IHPST, Paris, France

Thomas A. C. Reydon, Institute of Philosophy & CELLS, Leibniz Universität
Hannover, Hannover, Germany

Charles T. Wolfe, Département de Philosophie & ERRAPHIS, Université de
Toulouse Jean-Jaurès, Toulouse, France

Editorial Board Members

Marshall Abrams, University of Alabama, Birmingham, AL, USA

André Ariew, University of Missouri, Columbia, MO, USA

Domenico Bertoloni Meli, Indiana University, Bloomington, IN, USA

Richard Burian, Virginia Tech, Blacksburg, VA, USA

Minus van Baalen, Institut Biologie de l'Ecole Normale Supérieure, Paris, France

Pietro Corsi, University of Oxford, Oxford, UK

François Duchesneau, Université de Montréal, Montreal, QC, Canada

John Dupre, University of Exeter, Exeter, UK

Paul Farber, Oregon State University, Corvallis, OR, USA

Lisa Gannett, Saint Mary's University, Halifax, NS, Canada

Andy Gardner, University of St Andrews, St Andrews, UK

Jean Gayon, UFR de Philosophie, Université Paris 1 Panthéon-Sorbonne,
Paris, France

Guido Gigliani, University of Macerata, Civitanova Marche, Italy

Paul Griffiths, University of Sydney, Sydney, NSW, Australia

Thomas Heams, AgroParisTech, Paris, France

James G. Lennox, University of Pittsburgh, Pittsburgh, PA, USA

Annick Lesne, Sorbonne Université, Paris, France

Tim Lewens, University of Cambridge, Cambridge, UK

Edouard Machery, University of Pittsburgh, Pittsburgh, PA, USA

Alexandre Métraux, Archives Poincaré, Nancy, France

Hans Metz, Leiden University, Leiden, The Netherlands

Roberta L. Millstein, University of California, Davis, CA, USA

Staffan Müller-Wille, University of Cambridge, Cambridge, UK

François Munoz, Université Montpellier 2, Montpellier, France

Dominic Murphy, University of Sydney, Camperdown, NSW, Australia

Stuart A. Newman, New York Medical College, Valhalla, NY, USA
Frederik Nijhout, Duke University, Durham, NC, USA
Samir Okasha, University of Bristol, Bristol, UK
Susan Oyama, The City University of New York, New York, USA
Kevin Padian, University of California, Berkeley, CA, USA
David Queller, Washington University in St. Louis, St. Louis, MO, USA
Stephane Schmitt, Archives Poincaré, Nancy, France
Phillip Sloan, University of Notre Dame, Notre Dame, IN, USA
Jacqueline Sullivan, Western University, London, ON, Canada
Giuseppe Testa, University of Milan, Milan, Italy
J. Scott Turner, SUNY College of Environmental Science and Forestry,
Syracuse, NY, USA
Denis Walsh, University of Toronto, Toronto, ON, Canada
Marcel Weber, University of Geneva, Geneva, Switzerland

History, Philosophy and Theory of the Life Sciences is a space for dialogue between life scientists, philosophers and historians – welcoming both essays about the principles and domains of cutting-edge research in the life sciences, novel ways of tackling philosophical issues raised by the life sciences, as well as original research about the history of methods, ideas and tools, which constitute the genealogy of our current ways of understanding living phenomena.

The series is interested in receiving book proposals that • are aimed at academic audience of graduate level and up • combine historical and/or philosophical and/or theoretical studies with work from disciplines within the life sciences broadly conceived, including (but not limited to) the following areas: • Anatomy & Physiology • Behavioral Biology • Biochemistry • Bioscience and Society • Cell Biology • Conservation Biology • Developmental Biology • Ecology • Evolution & Diversity of Life • Genetics, Genomics & Disease • Genetics & Molecular Biology • Immunology & Medicine • Microbiology • Neuroscience • Plant Science • Psychiatry & Psychology • Structural Biology • Systems Biology • Systematic Biology, Phylogeny Reconstruction & Classification • Virology The series editors aim to make a first decision within 1 month of submission. In case of a positive first decision the work will be provisionally contracted: the final decision about publication will depend upon the result of the anonymous peer review of the complete manuscript. The series editors aim to have the work peer-reviewed within 3 months after submission of the complete manuscript. The series editors discourage the submission of manuscripts that contain reprints of previously published material and of manuscripts that are below 150 printed pages (75,000 words). For inquiries and submission of proposals prospective authors can contact one of the editors: Charles T. Wolfe: ctwolfe1@gmail.com Philippe Huneman: huneman@wanadoo.fr Thomas A.C. Reydon: reydon@ww.uni-hannover.de

João L. Cordovil • Gil Santos • Davide Vecchi
Editors

New Mechanism

Explanation, Emergence and Reduction

 Springer

Editors

João L. Cordovil
Centro de Filosofia das Ciências,
Departamento de História e Filosofia das
Ciências, Faculdade de Ciências
Universidade de Lisboa
Lisbon, Portugal

Gil Santos
Centro de Filosofia das Ciências,
Departamento de História e Filosofia das
Ciências, Faculdade de Ciências
Universidade de Lisboa
Lisbon, Portugal

Davide Vecchi
Centro de Filosofia das Ciências,
Departamento de História e Filosofia das
Ciências, Faculdade de Ciências
Universidade de Lisboa
Lisbon, Portugal



ISSN 2211-1948 ISSN 2211-1956 (electronic)
History, Philosophy and Theory of the Life Sciences
ISBN 978-3-031-46916-9 ISBN 978-3-031-46917-6 (eBook)
<https://doi.org/10.1007/978-3-031-46917-6>

© The Editor(s) (if applicable) and The Author(s) 2024. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable

Contents

1	A Framework for Mapping Mechanistic Perspectives	1
	João L. Cordovil, Gil Santos, and Davide Vecchi	
2	Different Types of Mechanistic Explanation and Their Ontological Implications.	9
	Beate Krickel	
2.1	Introduction	9
2.2	Constitutive Mechanistic Explanation: Minimal Characterization	11
2.3	The Functionalist View of Constitution and Constitutive Mechanistic Phenomena	14
2.4	The Behaving Entity View of Constitution and Constitutive Phenomena	17
2.5	Four Variants of Mechanistic Explanation	19
2.6	Constitutive Mechanistic Explanation: Different Ontological Implications	22
2.7	Conclusion	25
	References	26
3	The Metabolic Theory of Ecology as a Mechanistic Approach.	29
	Gonçalo Martins	
3.1	Introduction	29
3.2	The Metabolic Theory of Ecology	38
3.3	Virtues and Limitations of MTE	46
3.4	The Mechanistic Nature of MTE	52
3.5	Conclusion	57
	References	59

4	Causing and Composing Evolution: Lessons from Evo-Devo Mechanisms	61
	Cristina Villegas	
4.1	Introduction	61
4.2	Two Unusual Kinds of Biological Mechanisms	63
4.2.1	Mechanisms of Development.	64
4.2.2	Mechanisms of Evolution	67
4.3	<i>Evo</i> and <i>Devo</i> : The Mechanistic Composition of Variation.	70
4.4	Causing Phenotypic Change	73
4.4.1	Mechanistic Views of Innovation.	73
4.4.2	Innovation as an Evolutionary Mechanism	75
4.5	Concluding Remarks	79
	References.	80
5	Organisms Need Mechanisms; Mechanisms Need Organisms	85
	William Bechtel and Leonardo Bich	
5.1	Introduction	86
5.2	Constraints: A Revisionist Account of Mechanisms	88
5.3	Autonomy and the Closure of Constraints.	92
5.4	Control Mechanisms	95
5.5	Integrating Control Mechanisms	97
5.6	Conclusions	104
	References.	105
6	Searching for Protein Folding Mechanisms: On the Insoluble Contrast Between Thermodynamic and Kinetic Explanatory Approaches	109
	Gabriel Vallejos-Baccelliere and Davide Vecchi	
6.1	Introduction: What Is the Protein Folding Problem.	110
6.2	Brief Historical Overview of Folding Research.	112
6.3	Two Explanatory Approaches in Protein Folding Research	116
6.3.1	Thermodynamic and Kinetic Explanations	116
6.3.2	Mechanistic Credentials of Thermodynamic and Kinetic Explanations of Folding	118
6.4	Clashes Between Thermodynamic and Kinetic Approaches.	120
6.4.1	Micro Versus Macro Analyses	120
6.4.2	The Issue of Decomposition	125
6.5	What Kind of Explanations Are Thermodynamic Explanations of Folding?	127
6.5.1	Thermodynamic Explanations of Native state Stability	128
6.5.2	Thermodynamic Explanations of Folding Dynamics	129
6.6	Conclusion.	133
	References.	135

7	Mechanisms in Chemistry	139
	Robin Findlay Hendry	
	7.1 Introduction	139
	7.2 What Is a Chemical Reaction?	140
	7.3 What Is a Reaction Mechanism?	145
	7.4 Mechanisms and Reduction	152
	References	158
8	A Commentary on Robin Hendry's Views on Molecular Structure, Emergence and Chemical Bonding	161
	Eric Scerri	
	8.1 Introduction	162
	8.2 On Epistemological and Ontological Reduction	163
	8.3 Bonding	166
	8.4 Hendry's Contrast Between the Energetic and the Structural View of Bonding	168
	8.5 Quantum Mechanical Account of the Covalent Bond	170
	8.6 Are Bonds Real?	174
	8.7 Conclusions	175
	References	175
9	Fundamental Physics and (New-)Mechanistic Ontologies	179
	João L. Cordovil	
	9.1 Introduction	180
	9.2 Traditional Mechanical Philosophy in Physics	181
	9.2.1 Descartes	181
	9.2.2 Newton	182
	9.3 Contemporary Fundamental Physics and (New) Mechanical Philosophy	184
	9.3.1 Entanglement	184
	9.3.2 Still, the QM's Challenges	185
	9.4 Against the Universality Thesis of QM	186
	9.5 Conclusion	188
	References	188
10	Mechanistic Explanations in Physics: History, Scope, and Limits	191
	Brigitte Falkenburg	
	10.1 Introduction	191
	10.2 The Origin of Mechanistic Explanations	192
	10.2.1 The Tradition of Analysis and Synthesis	194
	10.2.2 Newton's Methodology	196
	10.3 Mechanistic Explanations Today	199
	10.3.1 The Recent Philosophical Definitions	200
	10.3.2 Causal Components and Their Dynamic Properties	202
	10.3.3 The "Atomistic" Constitution of Matter	203

10.4	Mechanistic Explanations in Neuroscience, and Their Limits.	205
10.5	Some Important Caveats	207
10.6	Summary and Conclusions	208
	Literature.	209
11	The Mechanisms of Emergence.	213
	Stuart Glennan	
11.1	Introduction: Mechanisms and Emergence	213
11.2	Mechanisms and their Varieties	215
11.3	Emergence as Mechanism-Dependence.	217
11.3.1	Producing versus Underlying and the Distinction between Diachronic and Synchronic Emergence	218
11.3.2	What Emerges: The Relata of Mechanism-Dependence Relations	219
11.4	Autonomy, Holism, and Novelty in Mechanistic Emergence.	222
11.4.1	Non-aggregativity	222
11.4.2	Externalism	223
11.4.3	Downward Causation.	224
11.4.4	Self-Organization.	225
11.4.5	Multiple Realization and Dynamical Autonomy	226
11.4.6	Transformation and Fusion	227
11.5	Conclusion: But is this Really Emergence?.	230
	References.	232
12	Emergence, Downward Causation, and Interlevel Integrative Explanations	235
	Gil Santos	
12.1	Introduction (to a Relational Ontological Approach)	235
12.2	Emergence	237
12.3	Downward Causation.	240
12.3.1	What Is a Whole?.	241
12.3.2	What Is the ‘Higher Level’ of an Integrated Whole?.	243
12.3.3	How Should We Conceptualize Downward Causation?.	244
12.3.4	How Does Downward Causation Work?	246
12.4	Interlevel Integrative Explanations	250
12.4.1	The Birth of a ‘New Mechanism’ and Its Integrative Explanation Models	250
12.4.2	Inter-theoretical Relations	254
12.4.3	Some Implications for a Neo-mechanistic Model of Explanation	260
	References.	261
	Index.	267

Contributors

William Bechtel Department of Philosophy, University of California, San Diego, La Jolla, CA, USA

Leonardo Bich IAS-Research Centre for Life, Mind and Society, Department of Philosophy, University of the Basque Country (UPV/EHU), Donostia-San Sebastian, Spain

João L. Cordovil Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Brigitte Falkenburg Technischen Universität Dortmund, Dortmund, Germany

Stuart Glennan Butler University, Indianapolis, IN, USA

Robin Findlay Hendry Department of Philosophy, Durham University, Durham, UK

Beate Krickel Technische Universität Berlin, Berlin, Germany

Gonçalo Martins Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Gil Santos Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Eric Scerri Department of Chemistry & Biochemistry, UCLA, Los Angeles, CA, USA

Gabriel Vallejos-Baccelliere Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

Daide Vecchi Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Cristina Villegas Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

Chapter 1

A Framework for Mapping Mechanistic Perspectives



João L. Cordovil, Gil Santos, and Davide Vecchi

This edited book is the outcome of a conference that was planned to take place in Lisbon at the Centro de Filosofia das Ciências (CFCUL) of the Faculdade de Ciências da Universidade de Lisboa. It was originally organized in the usual on-site form to which we were accustomed before the start of the Covid pandemic. The conference was postponed many times, with the hope to hold it on-site, to no avail. After many postponements, to our disappointment, the conference had unfortunately to be organized in a purely online form between 14th and 15th of October in 2021. The only advantage of all this is that we saved public money.

The eventual online conference was called *New Mechanism, Reduction and Emergence in Physics, Chemistry and Biology*. The participants were William Bechtel, Nancy Cartwright, Brigitte Falkenburg, Stuart Glennan, Robin Hendry, Alvaro Moreno and John Pemberton. Unfortunately, two of the talks (by Nancy Cartwright with John Pemberton and by Alvaro Moreno) did not result in a contribution to be published in the present volume. In the end, this book partially consists of a collection of articles based on some of the talks presented at the conference. Additionally, other contributions have been sought. It has not been easy at all to recruit other authors during the pandemic period. Our idea – already implicit in the conference title – was to seek contributions from research areas that have been somehow under-represented in the extant literature on new mechanism. We are therefore glad to have managed to enrol additional contributors, whose research encompasses several fields, including chemistry, biochemistry, developmental biology and ecology.

The idea for the conference originated from continuous conversations between its organizers, over many years, about the meaning of the qualification ‘new’ in what is today generally called “new mechanism” in philosophy of science. One significant aspect of the conversation concerned the potential *limits* new mechanism

J. L. Cordovil (✉) · G. Santos · D. Vecchi
Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências,
Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of
the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_1

faces when applied to areas of scientific research such as the quantum domain of physical reality, chemistry and biochemistry. Part of the rationale of the conference was thus to evaluate whether mechanistic analysis can be applied to sciences beyond those representing the original focus of new mechanism, particularly the molecular life sciences. An important caveat should be added at this juncture. Properly speaking, it is our contention that the advent of new mechanist perspectives in the twentieth century occurred simultaneously to the growing impact that cybernetics had, particularly on biology, after the late 1940s. By then, new mechanism was mainly seen as a way to overcome both old mechanist and neo-vitalist views. From the 1950s onwards, a variety of neo-mechanist approaches were developed (see Santos, Chap. 12 in this volume), the last of which being elaborated, from the 1990s, by the so-called “Chicago Mechanists”¹. The fundamental difference of these latter neo-mechanist views is that they were articulated as a new breed of philosophy of science, tailored to stand in opposition to the nomological theory of explanation and the theory-reduction model promoted by neo-positivism. These developments eventually engendered a series of questions concerning the domain of applicability of the mechanistic approach as well as the necessity of revising – or even expanding – the nature of mechanist analysis in order to account for recalcitrant natural phenomena.

In general terms, the book addresses the epistemological and ontological significance of new mechanism and, in particular, its relationship with the topics of neo-mechanist explanation, emergence and reduction in the physical, chemical and biological sciences. Several particular questions are targeted in this book. For example, how many different types of mechanistic explanation can we distinguish and accommodate (Krickel)? Can, or even should, new mechanism engage with historically antagonist biological traditions (Bechtel and Bich)? Can mechanistic analysis encompass (or even encroach on) seemingly non-mechanistic explanatory practices (e.g., stemming from thermodynamics) not even aiming to structurally decompose phenomena (Vallejos and Vecchi)? Does new mechanism fit the phenomena studied by contemporary sciences such as quantum mechanics (Cordovil and Falkenburg), chemistry (Hendry and Scerri), biochemistry (Vallejos and Vecchi), evolutionary developmental biology (Villegas) or ecology (Martins)? What can the new mechanistic position on the ongoing debate about the different notions of reduction and emergence, either in ontological or epistemological terms, be (Cordovil, Glennan, Hendry, Santos, and Scerri)? The ultimate aim of this book is to contribute to critically evaluate the scope of new mechanism in all the above respects.

In order to guide the reader, let us briefly elaborate on what should be considered the real *novelty* of any new mechanist perspective vis-à-vis the *old* seventeen-century mechanist philosophies. Paying attention to actual scientific practice is not a distinctive feature of new mechanism. Old mechanism was equally in tune with the

¹ Wimsatt, W. 2018. “Foreword”, In S. Glennan and P. Illari (Eds.), *Routledge handbook of mechanisms and mechanical philosophy* (pp. xiv–xvi). New York: Routledge.

scientific practice of its time. Moreover, attention to actual scientific practice is pervasive in serious history and philosophy of science, including the often-disparaged analyses of the neo-positivists. Given the different versions of the mechanistic perspective paving the history of philosophy and science, including the different *neo*-mechanist approaches emerging in the twentieth century, it might indeed be wondered whether there is anything distinctive constituting the theoretical core of any mechanistic view about the world. We would synthesize the core of old mechanistic philosophy in the following six features. **First**, mechanism postulates a pluralistic ontology of ontically discontinuous and discernible entities, even if spatially contiguous. In this sense, mechanism opposes absolute monism. **Second**, mechanism argues for the ontological and epistemological priority of *causal* relations and explanations. In this sense, mechanism opposes what would become radical empiricism and neo-positivism. **Third**, mechanism postulates the exclusive existence of *local* causal relations, either by direct contiguity or by propagation through a medium, therefore denying unmediated relations at a distance (which justifies Newton's problematic relation with his own theory of gravity as well as the contemporary problems mechanism encounters in accounting for phenomena such as quantum entanglement). **Fourth**, mechanism contends that, through their causal relations, entities form part-whole relations. In this sense, mechanism opposes mereological nihilism. **Fifth**, mechanism characterizes, as fundamental explanatory steps, the analytical tasks of decomposition and localization and the synthetic task of recomposition. **Sixth**, mechanism recognizes and highlights the existence of universal laws, causal or otherwise, expressible in mathematical terms.

While this characterization of the theoretical core of the old or original mechanistic view is not exhaustive, it remains useful to map the diversity of extant mechanistic approaches. More significantly, our characterization might be instrumental to identify the epistemological and ontological commitments of different versions of mechanism. Concerning the first and fifth features, mechanistic approaches might vary in recognizing the limitations of the analytic tasks of localization and decomposition; when failure of localization and/or decomposition rules, mechanistic analysis might be complemented by different analytic strategies (e.g., network analysis, dynamic systems theory, computational analysis, thermodynamic approaches). Whether these additional strategies might be considered mechanistic is under dispute, especially when they do not explicitly aim to open black boxes. Concerning the second feature, mechanistic approaches might vary in taking into consideration varieties of causal relations, with a basic opposition between those approaches privileging (or even merely countenancing) linear or additive relations and those encompassing non-linear and non-additive relations. Furthermore, mechanistic approaches might vary in relation to the nature of the kinds of changes that causal relations bring about in their relata (i.e., whether merely quantitative, qualitative and even substantial, that is, of kind). Mechanistic approaches might also vary in their implicit commitments to alternative ontologies, with the contrast between atomist/individualist essentialism and relationalism coming to the fore. Concerning the fourth feature, mechanistic approaches might vary in terms of endorsing or not an exclusive bottom-up or parts-to-whole ontological determination, without considering the

reversal form of partial and complementary, systemic co-determination. Correlatively, they also might vary in considering their ultimate aim as a whole-to-part reductive explanation. Relatedly, and concerning the fifth feature, mechanistic approaches might also vary concerning the necessity of including the additional synthetic task of (environmentally, contextually) “situating” when providing the explanation of any mechanism or system’s behaviour, thereby acknowledging the importance of not falling back into the classical isolated system view. Finally, mechanistic approaches might vary in terms of the relative epistemological role given to laws or law-like generalizations in the construction of science and scientific explanations, emphasizing instead the discovery of local mechanisms (this latter being a characteristic feature of the so-called “Chicago mechanism”). An additional distinctive feature would consist in defending the existence of different emergent laws and regional ontologies at different levels of organization or spatial-temporal scales. We would surmise that this minimal framework for mapping mechanistic perspectives might be helpful to navigate the ensuing contributions.

Let us finally introduce the themes of the book’s contributions and justify their sequential order. The first contribution, by Beate Krickel, deals with the nature of mechanistic explanation. As she argues, the assumption that there are just two kinds (i.e., *etiological* and *constitutive*) of mechanistic explanations is too narrow. Krickel therefore provides a quadripartite taxonomy, with two variants of etiological explanation—which she calls *output mechanistic explanations* and *input-output mechanistic explanations*—and two variants of constitutive explanation—which she calls *filler mechanistic explanations* and *dimensioned mechanistic explanations*. Krickel then delves on the differences between the two kinds of constitutive explanations, particularly in relation to the issues of reduction, mechanistic level and interlevel causation. The following eight contributions are focused on particular research fields. We have decided to organize them in a sequential order that some readers might considered topsy turvy, from ecology to physics, in descending order of system’s complexity (in the minimal sense of number and kinds of system’s parts and number and kinds of their interactions). We do not see any good reason to use the other sequential order. Gonçalo Martins focuses on mechanistic accounts in ecology, an area of research neglected in the extant mechanistic literature. Martins critically analyses the Metabolic Theory of Ecology. As its name suggests, this theory aims to account for population, community and ecosystem phenomena in terms of individual organisms’ metabolism. Martins acknowledges that the metabolic theory provides significant explanations of some phenomena at various levels of ecological organization. Nevertheless, he also argues that, first, the mechanistic nature of this approach needs further clarification and, secondly, that the metabolic theory is not able to completely elucidate the mechanistic basis of the ecological phenomena it explains. Cristina Villegas centres her analysis on evolutionary developmental biology (evo-devo), a field of research that features slightly more prominently than ecology in the extant mechanistic literature. This field is peculiar because practitioners often describe their explanations as mechanistic. It is also peculiar because, like

ecology, evolutionary issues are central. Villegas' aim is to provide a philosophical framework to make sense of the causal role of developmental processes in evolution. She therefore analyses the prospects and limits of a mechanistic view of evo-devo focusing on studies of homology and novelty. Finally, Villegas suggests a way to combine the mechanistic view of evo-devo with the population-level analysis of classical approaches to evolution. Next in the sequence is the contribution by William Bechtel and Leonardo Bich. This paper prolongs the effort to expand the classical version of Chicago's new mechanism by promoting a constructive engagement with the autonomy tradition centred on organismal self-maintenance. Bechtel and Bich argue that a natural linkage between these two traditions is given by the fact that self-maintenance relies on mechanisms. What the autonomy tradition adds to this picture is the notion of control, which in its turn implies, Bechtel and Bich argue, characterizing mechanisms as sets of constraints on the flow of free energy. The relationship between control and controlled mechanisms is, they finally argue, heterarchical. In their contribution, Vallejos and Vecchi analyse two different biochemical approaches to the protein folding problem: kinetic approaches are intuitively mechanistic, aiming to reconstruct folding pathways in terms of structural considerations; thermodynamic approaches instead focus on energetic considerations, neglecting structural changes. After briefly illustrating the origin of these alternative approaches, Vallejos and Vecchi characterise their contrasting epistemological and ontological commitments. They then critically analyse in what sense thermodynamic explanations of folding might be said to be mechanistic or causal. The underlying issue – implicit in Bechtel and Bich's as well as Hendry's and Scerri's contributions – concerns the possibility of meaningfully combining thermodynamic and mechanistic analyses. Robin Hendry centres his analysis on the nature of reaction mechanisms in chemistry. Mechanistic explanations of chemical reactions are – as Vallejos and Vecchi relate in the case of biochemistry – kinetic in nature. These explanations aim to identify significant chemical pathways, decomposing them into a series of steps involving structural modifications such as the breaking and making of bonds. The problem Hendry addresses is whether the establishment of a reaction mechanism vindicates the reduction of chemistry to physics. Hendry argues that, while in a sense this might be considered the case (chemical processes basically involve transfers of conserved quantities), in another sense, arguably more significant, reduction is not vindicated. Eric Scerri's contribution aims to critically evaluate some of Hendry's arguments in support of emergence and downward causation in chemistry as well as on the nature of the chemical bond. In the first sense, Scerri argues that alternative explanations (e.g., based on the notion of quantum decoherence) of the compositional identity but structural difference of isomers make emergence and downward causation redundant. In the second sense, Scerri points again at the structural vs. thermodynamic contrast underlying the chemical sciences. In particular, he argues that, while it is true that chemists view bonding in a more realistic fashion while physicists consider bonding in more abstract energetic terms, such differences in scientific practice do not substantiate specific views about the ontological status of bonding. João L. Cordovil's contribution argues that the challenges posed by Quantum Mechanics to mechanism are not

substantially new, since there has always been a problematic relationship between mechanical philosophy and fundamental physics throughout the history of physics. Despite this, mechanism always prevailed. According to Cordovil, although fundamental physics may not be compatible with new mechanism, this incompatibility can only be considered as a fundamental problem if we uphold the micro-physicalist assumption concerning the universal character of quantum mechanics. Cordovil thus suggests that, rather than trying to find an answer to this problem in the quantum decoherence hypothesis, it would be better to consider the ways in which the classical physical domain might have emerged from the quantum domain of physical reality. In her contribution, Brigitte Falkenburg argues that, notwithstanding the scientific revolutions of the twentieth century, mechanistic approaches continue to be based on the traditional method of analysis and synthesis and, therefore, on the assumption that all higher-level phenomena are to be explained in terms of lower-level parts' properties, their interactions, and some composition rules. Nevertheless, quantum fields, as well as higher-level phenomena (e.g., chemical, biochemical, and biological) pose challenges to the mechanistic approach. Thus Falkenburg asks: is it just a mere *façon de parler* to talk of mechanisms underlying such phenomena? In particular, Falkenburg points out, no mechanism is known that might explain how the brain produces the conscious human mind. The last two chapters focus on the topic of emergence and its relationship with the mechanistic approach. In his contribution, Stuart Glennan aims to show that the opposition between mechanism and emergence is essentially based on a misunderstanding and that the core features of emergent phenomena (dependence, autonomy, holism and novelty) can be explicated in mechanistic terms. Indeed, according to Glennan, if there are naturalistic processes of emergence there must be mechanisms responsible for their existence. Furthermore, the mechanistic view allows the possibility of classifying different kinds of emergent phenomena in terms of the particular features of the mechanisms generating them. For example, the distinction between mechanisms that produce phenomena vs. mechanisms that underlie phenomena provides an analysis of the distinction between diachronic and synchronic emergence, and various interpretations of novelty, holism and autonomy can then be shown to arise from different kinds of mechanistic organization. Gil Santos' contribution proposes a dynamic relational account of both systemic emergence and downward causation. In particular, Santos argues for a relational-transformational notion of emergence and a structural-relational account of downward causation in terms of both its transformational and conditioning effects. According to Santos, it is the objective existence of systemic emergence and downward-structural causation that ultimately justifies the in-principle failure of any form of micro-determinism and micro-reductionism, and that at the same time most strongly requires the use of interlevel integrative forms of explanation. Furthermore, according to the author, it is here that one may find the real ontological and epistemological novelty of any neo-mechanistic view in comparison to the old seventeenth-century mechanistic philosophies. We wish you an enjoyable read.

Acknowledgements The editors acknowledge the financial support of the FCT – Fundação para a Ciência e a Tecnologia (Grants N. UIDB/00678/2020 and UIDP/00678/2020). In particular, we must acknowledge the financial support of the FCT – Fundação para a Ciência e a Tecnologia (R&D Project Grant PTDC/FER-HFC/30665/2017 “Emergence in the Natural Sciences: Towards a New Paradigm”) for this book’s Open Access publication. We would like to especially thank our Faculty of Letters university colleague David Yates, principal investigator of the above mentioned “Emergence in the Natural Sciences: Towards a New Paradigm” project, for making the Open Access publication possible. João L Cordovil acknowledges the financial support of FCT, ‘Fundação para a Ciência e a Tecnologia, I.P.’ (Stimulus of Scientific Employment, Norma Transitória: DL57/2016/CP1479/CT0065). Gil Santos acknowledges the financial support of FCT, ‘Fundação para a Ciência e a Tecnologia, I.P.’ (Stimulus of Scientific Employment, Individual Support 2017: CEECIND/03316/2017). Davide Vecchi acknowledges the financial support of the FCT— Fundação para a Ciência e a Tecnologia (DL57/2016/CP1479/CT007).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 2

Different Types of Mechanistic Explanation and Their Ontological Implications



Beate Krickel

Abstract One assumption of the new mechanistic approach is that there are two kinds of mechanistic explanations: *etiological* and *constitutive* ones. While the former explain phenomena in terms of their preceding causes, the latter are supposed to refer to mechanisms that *constitute* phenomena. Based on arguments by Kaiser and Krickel (Br J Philos Sci 68(3):745–779, 2017) and Krickel (The mechanical world, vol. 13, Springer International Publishing. <https://doi.org/10.1007/978-3-030-03629-4>, 2018), I will show that this view is too narrow. Indeed, three different types of explanation are usually subsumed under the label “constitutive explanation”. However, one of those types of explanation is not a version of constitutive explanation. Rather it is a variant of etiological explanation. As a result, I will show that there are four types of mechanistic explanation, two variants of etiological explanation—which I will call *output mechanistic explanations* and *input-output mechanistic explanations*—and two variants of constitutive explanation—which I will call *filler mechanistic explanations* and *dimensioned mechanistic explanations*. Keeping these apart is crucial as they come with different ontological implications. An evaluation of the mechanistic approach regarding its stance on reduction, levels, and interlevel causation crucially depends on which notion of mechanistic explanation one has in mind.

Keywords Mechanisms · Mechanistic explanation · Causal explanation · Functionalism · Dimensioned approach

2.1 Introduction

In the new mechanistic literature, authors often distinguish between two different types of mechanistic explanation: *etiological* mechanistic explanation, in which a phenomenon is explained by its preceding causes, and *constitutive* mechanistic

B. Krickel (✉)
Technische Universität Berlin, Berlin, Germany
e-mail: beate.krickel@tu-berlin.de

explanations, which refer to mechanisms that “constitute” the phenomenon-to-be-explained (Craver, 2007a, b; Salmon, 1984a, b). The distinction between these two types of explanation was first introduced by Wesley Salmon (1984a, b). Constitutive explanations, according to Salmon, “account for a given phenomenon by providing a causal analysis of the phenomenon itself” (1984a, p. 297), while an etiological explanation “tells the causal story leading up to its occurrence” (ibid.). Salmon’s focus was on etiological explanation. Carl Craver (2007b) prominently highlighted the relevance of constitutive explanations for the life sciences. He discusses various examples of constitutive mechanistic explanations, such as the explanation of the action potential and spatial memory. Most philosophers of the life sciences and the cognitive sciences now agree that constitutive mechanistic explanation is ubiquitous.

While the notion of causation is extensively discussed in philosophy of science as well as metaphysics, the debate on mechanistic constitution is quite young and there are still many open questions such as whether mechanistic constitution can be spelled out in terms of interventionism, whether approaches to constitution should be singularistic or generalistic, which role time plays in constitution, what the relations of constitution are, and how constitution differs from causation (Baetu, 2012; Baumgartner & Gebharter, 2016; Couch, 2011; Craver, 2007a, b; Craver et al., 2021; Fagan, 2012; Gillett, 2013; Harbecke, 2010; Harinen, 2018; Kästner, 2017; Kirchhoff, 2017; Kistler, 2009; Krickel, 2018b; Romero, 2015; Weinberger, 2019). The new mechanistic account of constitutive mechanistic explanation is not only supposed to capture a common and important explanatory practice in the life sciences. The nature of constitutive mechanistic explanation, according to the new mechanists, has several implications for various ontological questions as well. For example, whether or in which sense mechanistic explanation is reductive, what levels of nature are, and whether there can be causal relationships between these levels directly depends on how the notion of constitutive mechanistic explanation is understood. Thus, the details of the account of constitutive mechanistic explanations are crucial for the evaluation of the ontological implications of the new mechanistic account.

This chapter has two goals: First, I will use ideas developed by Krickel (2018a) and Kaiser and Krickel (2017) to show that there are three different interpretations of what constitutive mechanistic explanation amounts to. I will use these considerations to argue that there are in fact four variants of mechanistic explanation, not just two. Second, I will describe the different ontological implications of the different versions of constitutive mechanistic explanation and outline the different pictures of reduction, levels, and interlevel causation that they suggest.

The paper proceeds as follows: in Sect. 2.2, I will present the general features that are commonly attributed to constitutive mechanistic explanation. In Sects. 2.3 and 2.4, I will summarize the two views of mechanistic constitution and mechanistic phenomena presented by Krickel (2018a)—the *functionalist view of constitutive mechanistic phenomena* and the *behaving entity view of constitutive mechanistic phenomena*. I will recap three possible interpretations of the functionalist view as presented in Krickel (2018a) and Kaiser and Krickel (2017). In Sect. 2.5, I will show that the different views on mechanistic constitution and constitutive

phenomena suggest that there are indeed *four* different types of mechanistic explanation—each of which describes a common explanatory practice in the life sciences. In Sect. 2.6, I will outline the different ontological consequences of the different types of mechanistic explanation.

2.2 Constitutive Mechanistic Explanation: Minimal Characterization

What is constitutive mechanistic explanation? In this section, I will provide a minimal characterization in terms of a list of criteria of adequacy that an approach to constitutive mechanistic explanation must satisfy.

It is commonly assumed that mechanistic constitutive explanation is a type of mechanistic explanation where a phenomenon is explained by its *underlying* mechanism. The relation between the phenomenon and the mechanism is usually called “mechanistic constitution” and the relation between a component of the mechanism and the phenomenon “constitutive relevance”. To get a grasp of what mechanistic constitutive explanation is, let us briefly summarize what the mechanists take mechanisms to be.

The nature of mechanisms has gained a lot of attention in the mechanistic literature. Here is what Glennan (2017) calls the *minimal characterization* (MC) of a mechanism:

(MC)	A mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon. (Glennan, 2017, 17; for similar formulations see Craver (2007b) and Illari and Williamson (2012)).
------	---

This characterization applies to mechanisms in general, i.e., to those that are referred to in etiological *and* those that are referred to in constitutive mechanistic explanations (Craver, 2007b, p. 22). To illustrate how this characterization applies to constitutive mechanistic explanation, consider the example of the action potential mechanism (Craver, 2007b, Chap. 4). This mechanism consists of various entities such as different types of ions and ion-channels. These ions and ion-channels are engaged in different activities such as opening, closing, or diffusing. Furthermore, these entities and activities are organized in various ways. For example, the ion-channels are spatially organized along the axon, and the location of the ions outside and inside the axon is crucial for the working of the mechanism. The different activities are temporally organized. For example, the temporal order of the opening and closing of the ion-channels is crucial for the action potential to occur. Furthermore, the components of mechanisms are what Craver calls “actively” organized (Craver, 2007b, p. 136). They interact in various ways. For example, the ions diffuse through the ion-channels.

The central idea of a constitutive mechanistic explanation is illustrated in the well-known *Craver-diagram* (see Fig. 2.1). This diagram has become popular in the

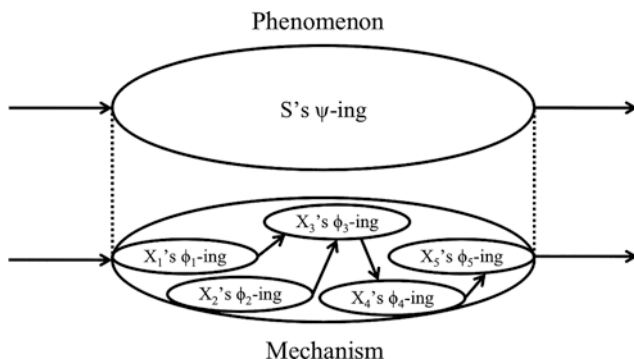


Fig. 2.1 The *Craver diagram*: Illustration of a mechanism constituting a phenomenon. (Adapted from Craver (2007b, p. 6))

new mechanistic literature and most authors take it to be an adequate illustration of a mechanism that constitutes a phenomenon.¹

In Fig. 2.1, the mechanism is located at the bottom (the different Xs stand for the entities and the different ϕ -ings stand for their activities; the arrows inside the lower big circle indicate the interactions between the entities and activities). The phenomenon consists of a system (“S”) that is engaged in a certain behavior (“ ψ -ing”). Referring to the phenomenon by “S’s ψ -ing” or “S ψ -ing” has become a common convention in the new mechanistic literature.

The minimal characterization of a mechanism presented above seems to apply well to the example of the action potential mechanism. But what renders the explanation of the action potential a *constitutive* mechanistic explanation? As indicated in the Craver-diagram, one core feature of mechanistic constitution is that it relates a mechanism and a phenomenon that occur at the same time. Thus, in constitutive mechanistic explanations, the mechanism and the phenomenon do not stand in a causal relationship as a central assumption about causation is that causes and effects occur at different times (Lewis, 1973).

Further features of the relation between mechanisms and phenomena can be inferred from Craver’s approach to *constitutive relevance* (Craver, 2007a, b) that has become the standard account of constitutive mechanistic explanation (Casini et al., 2011; Gillett, 2013; Illari & Williamson, 2011; Irvine, 2013; Kaplan, 2012; Levy, 2009; van Eck & Looren de Jong, 2016; Zednik, 2015). Craver argues that in constitutive mechanistic explanations the components of a mechanism are not

¹Note that the constitution-relation is taken to hold between the mechanism and the phenomenon, not between the mechanism and its components. Mechanisms *are* (cf. Craver 2007, 5) sets of or *consist of* (Glennan, 2017, 17) entities and activities organized such that they bring about/exhibit/are responsible for/cause/constitute the phenomenon.

causally but *constitutively* relevant for the phenomenon. According to Craver's account, a component X's ϕ -ing is constitutively relevant for a phenomenon S's ψ -ing if²:

- (i) X's ϕ -ing is a spatiotemporal part of S's ψ -ing, and
- (ii) X's ϕ -ing and S's ψ -ing are mutually manipulable (Craver, 2007b)

Condition (i) specifies that the entity-component must be a spatiotemporal part of the system whose behavior is to be explained, and the activity-component has to be executed during the system's ψ -ing. Condition (ii) is spelled out in terms of interventionism (Woodward, 2003, 2015): X's ϕ -ing and S's ψ -ing are mutually manipulable if and only if it is possible to ideally intervene into X's ϕ -ing and thereby change S's ψ -ing, and it is possible to ideally intervene into S's ψ -ing and thereby change X's ϕ -ing. An intuitive understanding of interventions suffices for present purposes. Many authors argue that the mutual manipulability account and interventionism are incompatible (Baumgartner & Gebharder, 2016; Kästner, 2017; Leuridan, 2012; Romero, 2015). Therefore, it remains controversial how to understand the claim that phenomena and mechanisms mutually depend on each other (note that promising attempts have been made to save the combination between constitutive explanation and interventionism; see (Baumgartner & Gebharder, 2016; Craver et al., 2021; Krickel, 2018b; Romero, 2015)). For the sake of argument, I will formulate the mutual manipulability requirement as a general *mutual dependency* requirement in the sense of "If the phenomenon had been different, the mechanism would have been different; and if the mechanism had been different, the phenomenon would have been different", while this is supposed to remain silent with regard to how this mutual dependency relation is to be spelled out.

In summary, there are at least six characteristics that are commonly attributed to constitutive mechanistic explanation:

1. *Mechanism*: Mechanisms are entities and activities organized such that they are responsible for a phenomenon.
2. *Phenomenon*: The phenomenon consists of a system S showing a behavior ψ -ing.
3. *Non-Causal*: The mechanism does not cause the phenomenon.
4. *Temporal Synchrony*: The phenomenon and the mechanism occur at the same time.

²Some authors, such as Baumgartner and Gebharder (2016), interpret the mutual manipulability account as providing necessary and sufficient criteria for constitutive relevance. This interpretation is supported by Craver's summary of his account in his 2007-book: "In sum, I conjecture that to establish that X's ϕ -ing is relevant to S's ψ -ing it is sufficient that one be able to manipulate S's ψ -ing by intervening to change X's ϕ -ing (by stimulating or inhibiting) and that one be able to manipulate X's ϕ -ing by manipulating S's ψ -ing. To establish that a component is irrelevant, it is sufficient to show that one cannot manipulate S's ψ -ing by intervening to change X's ϕ -ing and that one cannot manipulate X's ϕ -ing by manipulating S's ψ -ing" (Craver, 2007b, p. 159). However, in later works, Craver states that the mutual manipulability account is meant to provide sufficient conditions only (see Craver et al., 2021, especially fn. 7).

5. *Spatiotemporal Part-Whole Relation*: The mechanism's components are spatio-temporal parts of the phenomenon.
6. *Mutual dependency*: If the phenomenon had been different, the mechanism would have been different; and if the mechanism had been different, the phenomenon would have been different.

There seems to be a general agreement that constitutive mechanistic explanation has these six features. Still, as Krickel (2018a, Chap. 6) and Kaiser and Krickel (2017) show, there are different ways of how to account for these features. In the following two sections, I will summarize Krickel's and Kaiser and Krickel's considerations concerning constitutive mechanistic phenomena and mechanistic constitution that suggest that there are different interpretations of what constitutive mechanistic explanation amounts to.

2.3 The Functionalist View of Constitution and Constitutive Mechanistic Phenomena

Krickel (2018a) and Kaiser and Krickel (2017) discuss the nature of what they call "constitutive mechanistic phenomena", i.e., phenomena that form the explananda of constitutive mechanistic explanation. One possible interpretation of the nature of constitutive phenomena is what Krickel (2018a) calls the *functionalist view of constitutive mechanistic phenomena*. According to this view, the system whose behavior is to be explained is the mechanism itself (see Krickel (2018a, Chap. 6)). Hence, the thing that is mechanistically constituted (the "constituee") is the behaving mechanism. As Krickel (2018a) shows, this idea underlies many discussions in the new mechanistic literature (Bechtel & Abrahamsen, 2005, p. 426; Craver, 2007b, pp. 6–7, 128; Fagan, 2012, p. 467; Fazekas & Kertész, 2011; Illari & Williamson, 2012). In this picture, the behavior that is to be explained is commonly characterized in terms of a complex input-output relation or a causal role. According to the functionalist view, these inputs and outputs are connected by the mechanism (Baetu, 2012; Bechtel, 2008, pp. 201–202; Craver, 2007b, pp. 146, 214; Craver et al., 2021; Fazekas & Kertész, 2011; Kuorikoski, 2012, pp. 146, 375; Soom, 2012). The functionalist view of constitutive mechanistic phenomena can be summarized as follows:

- (i) The constituee is a *behaving mechanism*, and
- (ii) the behavior is characterized in terms of *inputs* and *outputs* of the mechanism.

According to the functionalist view of constitutive mechanistic phenomena all there is, is a mechanism that plays a certain causal role, that connects certain inputs with certain outputs. To capture this idea, we can modify the Craver-diagram (presented in Fig. 2.1). Figure 2.2 illustrates what a mechanism constituting a phenomenon looks like according to the functionalist view (a similar picture can be found in (Bechtel, 2017) and Fazekas (2022)).

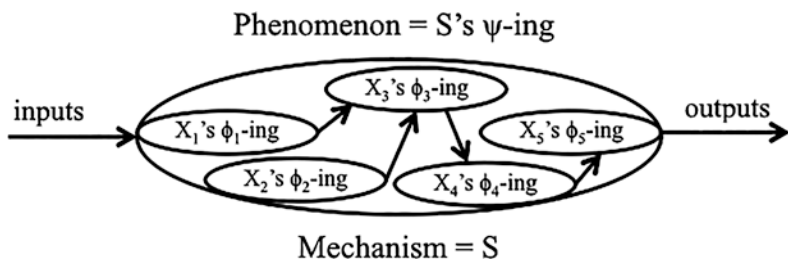


Fig. 2.2 Illustration of a mechanism constituting a phenomenon according to the functionalist view (my illustration based on Bechtel, 2017, fig. 3; and Fazekas, 2022, fig. 2)

There are three possible interpretations of this picture. One interpretation can be found in Kaiser and Krickel (2017). It may be called the *input-output functionalist view* of constitutive mechanistic explanation. According to this view, the phenomenon to be explained *just is the inputs plus the outputs* of a mechanism. In that sense, the phenomenon is a set of causes and effects. Changing the phenomenon is to change the input or the output. Under this interpretation, Fig. 2.2 must be interpreted accordingly: the phenomenon/S's ψ -ing is not the circle in the middle but the inputs and the outputs.

The input-output functionalist view, however, is not a valid interpretation of mechanistic constitution. First, it violates *Non-Causal*: if the phenomenon is the inputs plus the outputs of the mechanism, the two stand in a causal relation. Indeed, the only difference to standard etiological mechanistic explanation is that one cause of the phenomenon is explicitly picked out as the input to the mechanism—which seems to be a difference only in labelling or perspective and not a substantial metaphysical difference. Furthermore, this account violates *Temporal Synchrony*: the mechanism and the phenomenon do not occur at the same time. Rather, the phenomenon occurs before and after the mechanism occurs. And it violates *Spatiotemporal Part-Whole Relation* as the mechanism's components are not parts of the inputs and outputs, and thus, they are not parts of the phenomenon. Hence, the input-output functionalist view should not be considered an account of mechanistic constitution and constitutive mechanistic explanation. Still, as I will argue in Sect. 2.5, the input-output functionalist view gives rise to a valid account of mechanistic explanation (just not *constitutive* mechanistic explanation).

The second interpretation of the functionalist view is in terms of *role functionalism* (see Krickel (2018a)). According to mechanistic role functionalists, the phenomenon is a *causal role*, and the relation of mechanistic constitution is that of *causal role-playing*. However, again, the role functionalist interpretation cannot be regarded as an account of mechanistic constitution because it does not satisfy the criteria listed in Sect. 2.2. According to role functionalism, phenomena are abstract entities (functional second-order properties). Abstract entities are not located in space and time, and they do not have spatiotemporal parts. Hence, the role functionalist view cannot account for the fifth criterion (*Spatial Part-Whole Relation*).

What is more: if phenomena are causal roles in the role functionalist sense, there cannot be mechanistic explanations of them as this view would either be inconsistent, or it would collapse into the realizer functionalist view (see below). Causal roles in the role functionalist pictures surely have mechanistic realizers (if physicalism is presupposed). But this is an ontological, not an explanatory statement. The whole idea of role functionalism is to provide an explanation that abstracts away from the details of the realizer. The explanatory power comes from the fact that there are different mechanisms that despite their differences play the same causal role. Adding information about the realizer will not improve the explanation as the explanatory generalization is said to be found at the level of the abstract causal role. This reasoning is known from the discussion of non-reductive physicalism and multiple realization. Irrespective of whether role functionalism is a convincing ontological theory, if there are explanations of causal roles in the sense of role functionalism, these will not be constitutive mechanistic ones but, rather, functional explanations. For example, the question “Why do all these chemical substances function as neurotransmitters?” will be answered by citing the causal role that characterize neurotransmitters and by stating that all these chemical substances execute this causal role. If one takes the mechanistic details of the realizer to be explanatory relevant of the causal role, indeed, one is advocating the realizer functionalist interpretation (see below). Thus, the role functionalist view is incompatible with the mechanistic account in general.

A more promising interpretation of the functionalist view that indeed gives rise to a version of mechanistic explanation is the *realizer functionalist view of constitutive mechanistic phenomena*.³ According to this interpretation, in a first step the phenomenon is characterized in terms of an input-output relation, or a causal role. In a second step, the mechanism that plays this causal role is identified. Then, the phenomenon is identified with the mechanism. According to realizer functionalism, the phenomenon just is the mechanism under a functional description. Mechanistic constitution, here, is identity.

The realizer functionalist view has gained some prominence among defenders of so-called *causal betweenness accounts of constitutive relevance*. A formulation of this account can be found in a recent article by Craver et al. (2021).

Here, ψ -ing is represented as a process beginning with an input, ψ_{in} , and terminating with an output, ψ_{out} . Between these temporal endpoints, and a mechanistic level down, is a temporally sequenced causal chain of events, involving the X_i and their various ϕ_i . (...) [T]he problem of constitutive relevance is that of identifying the components of the process bridging ψ_{in} and ψ_{out} : What lies on the causal path(s) between these phenomenon-defining endpoints? The higher-level activity, ψ -ing, just is an organized collection of $X_i \phi_i$ -ing. (Craver et al., 2021, p. 8812)

The *realizer functionalist view* accounts for all features of constitutive mechanistic explanation listed in Sect. 2.2. The phenomenon, according to this interpretation, just is the mechanism under a functional description. Hence, the relation between

³For a general explanation of the common distinction between role and realizer functionalism see Levin (2023).

the mechanism and the phenomenon is not causation but identity. The mechanism cannot cause itself. Hence, the realizer functionalist view accounts for *Non-Causal*. Indeed, the realizer functionalist interpretation of mechanistic constitution accounts for all further features on the list—for a rather trivial reason: since the phenomenon just *is* the mechanism, trivially, the former occurs at the same time as the latter, parts of the latter are parts of the former, and changing one leads to changes in the other, and *vice versa*.

Other defenders of causal betweenness accounts of constitutive relevance, however, reject the third step, i.e., the identification of the phenomenon with the mechanism. Totte Harinen (2018), for example, is skeptical.

Is S's ψ -ing something *over and above* of the organized ϕ -ings of all of the Xs passing the mutual manipulability test, that is, $X_{1,\dots,n}$'s $\phi_{1,\dots,n}$ -ing? Most philosophers and scientist would probably agree that there is some sense in which S's ψ -ing is indeed more than just the sum of the ϕ -ings of its Xs, but that the relation between the two should not be that of spooky, materialistically inexplicable *emergence*. At the same time, many would not want to *identify* S's ψ -ing with $X_{1,\dots,n}$'s $\phi_{1,\dots,n}$ -ing, and so there is a market for an intermediate type of interlevel relation. (Harinen, 2018, 40)

Defenders of causal betweenness accounts of constitutive relevance who are not convinced by the identity claim, such as Harinen, argue for (a version of) what Krickel (2018a) calls the *behaving entity view of constitution*—which I summarize and discuss in the next section.

2.4 The Behaving Entity View of Constitution and Constitutive Phenomena

A further possible interpretation of constitutive mechanistic phenomena, according to Krickel (2018a), is the *behaving entity view*. This view is characterized by two claims:

- (i) the constitutee is the behavior of an object or system that contains the mechanism, and
- (ii) this behavior is an activity of the object or system.

According to Krickel, this view can be found in Craver's discussion of spatial memory (Craver, 2007b), in Stuart Glennan's work who talks about mechanisms located inside watches, cells, organisms, or toilets (Glennan, 1996, 2002), as well as in Carl Gillett's interpretation of mechanistic constitution in terms of dimensioned realization (Gillett, 2013, pp. 327–328). According to the behaving entity view, the relation between a mechanism and a phenomenon in constitutive mechanistic explanations can be illustrated as shown in Fig. 2.3.

As Fig. 2.3 shows, the analysis of the mechanism according to the behaving entity view and the functionalist view are identical (in both cases mechanisms are entities (the Xs) and activities (the ϕ -ings) in a certain organization). The difference between the two views is that, according to the behaving entity view, the

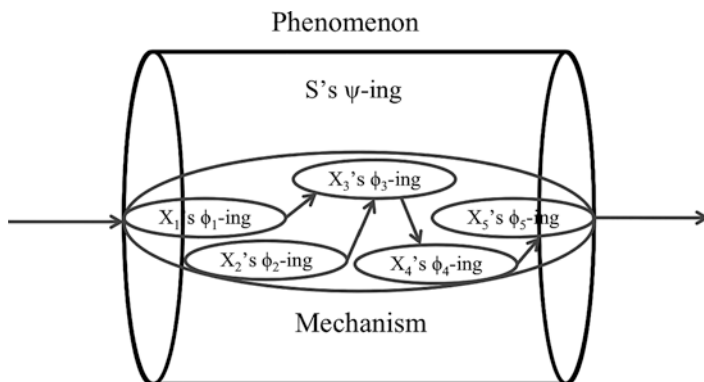


Fig. 2.3 Illustration of a mechanism constituting a phenomenon according to the behaving entity view

phenomenon is not a behavior of the mechanism. Rather, the phenomenon (S 's ψ -ing) *contains the mechanism* (in Fig. 2.3, the bigger tube that represents the phenomenon contains the circle that stands for the mechanism). The containment relation is a spatial and a temporal one: the mechanism's entity-components (the X s) are spatiotemporal parts of S and the different activity-components of the mechanism (the ϕ -ings) occur during S 's ψ -ing.

The behaving entity view accounts for all features listed in Sect. 2.2. Consider the explanation of muscle contraction. According to the behaving entity view, the phenomenon is *the muscle contracting* (an object showing a certain behavior). The mechanism for muscle contraction consists of various entities, such as actin and myosin filaments and ATP, and activities, such as binding, detaching, energizing and rotating in a certain organization. According to the behaving entity view, the relation between the mechanism for muscle contraction and the muscle's contracting cannot be causal. The reason is that the muscle's behavior and the mechanism responsible for this behavior occur at the same time. Furthermore, the muscle's behavior depends on the behaviors of the filaments. Hence, the phenomenon and the mechanism are not wholly distinct. Distinctness is required for causation. Hence, the behaving entity view accounts for the (*Non-Causal*) requirement. This shows that (*Temporal Synchrony*) is accounted for by the behaving entity view as well. The muscle that contains the mechanism for muscle contraction shows the contracting-behavior only during the occurrence of the contracting-mechanism. Furthermore, the mechanism's components, the actin and myosin filaments and their behaviors, are spatiotemporal parts of the muscle. Hence (*Spatial Part-Whole Relation*) is accounted for by the behaving entity view.

As mentioned in Sect. 2.2, it is controversial how to account for the last feature (*Mutual Dependency*). Different authors have argued that the combination of constitutive relationships and interventionism (that provides the framework for mutual manipulability) is problematic. One reason is that interventionism is supposed to be an approach to causation while constitution is supposed to be a non-causal dependency-relation. At least *prima facie* interventionism is inapplicable to

non-causal dependency relations. I will not try to solve this issue here. The applicability of the interventionist framework to mechanistic constitution is a topic of an ongoing debate (Baumgartner & Gebharder, 2016; Craver et al., 2021; Harinen, 2018; Kästner, 2017; Krickel, 2018b; Romero, 2015; Woodward, 2015). Note that the causal betweenness interpretation of constitutive relevance mentioned in Sect. 2.3 that has been put forward as a solution to the problem can be combined with the behaving entity view as well. Along the lines of the causal betweenness account, the mutual dependency between the behaving entity (i.e., the phenomenon) and the mechanism can be interpreted as the causal influence of the input that triggers the entity’s behavior and the mechanism’s components—the top (the top-down intervention) and the causal influence of the mechanism’s components and the output that is produced by the behaving entity (the bottom-up intervention).

The behaving entity view sheds a different light on constitutive explanation than the functionalist view. First, the behaving entity view is not committed to an identity-claim. On that view, the mechanism is a *proper* spatiotemporal part of the system whose behavior is to be explained. Not all parts of the system are involved in the mechanism (this is why the distinction between relevant and irrelevant parts is so crucial for the new mechanistic approach). For example, the mechanism that explains muscle contraction does not involve all parts of the muscle. Hence, the contracting muscle (the phenomenon) cannot be identical to the mechanism that is responsible for the muscle’s contracting. Similarly, the mechanism that explains the moving of a car does not contain the car’s doors. Hence, the moving car (the phenomenon) cannot be identical to the driving mechanism. Furthermore, following Gillett, the entities that are related constitutively in the way captured by the behaving entity view are what he calls “qualitatively distinct”, they have substantially different features and can enter different causal interactions (2010, p. 172). This qualitative distinctness between the related blocks an identity-claim.

2.5 Four Variants of Mechanistic Explanation

The foregoing discussion has shown that there are four variants of mechanistic explanation. Table 2.1 provides an overview of these four variants.

As shown in Table 2.1, *output mechanistic explanation* is what is standardly called “etiological mechanistic explanation”. I chose a different label because, as shown in Sect. 2.3, there is a further version of etiological mechanistic explanation:

Table 2.1 Overview of the four variants of mechanistic explanation

Etiological		Constitutive	
(1) <i>output mechanistic explanations</i>	(2) <i>input-output mechanistic explanations</i>	(3) <i>filler mechanistic explanations</i>	(4) <i>dimensioned mechanistic explanation</i>
Former etiological explanation	Input-output functionalist view	Realizer functionalist view	Behaving entity view

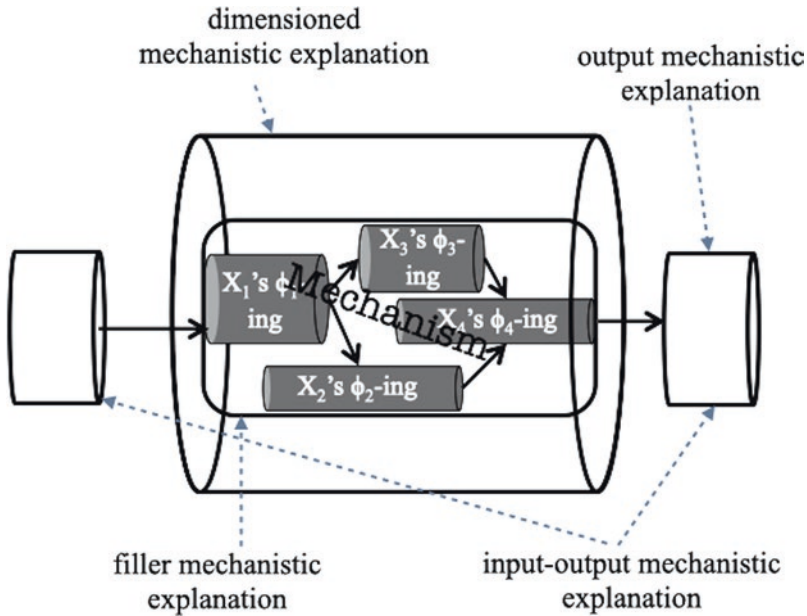


Fig. 2.4 The four types of mechanistic explanation grasp different aspects of the same going-on in the world. The grey tubes in the middle represent the mechanism. The arrows point to the different phenomena that are explained by the mechanism

input-output explanation. The latter two variants—*filler mechanistic explanation* and *dimensioned mechanistic explanation*—can plausibly be described as versions of constitutive explanation. The explanantia of all four types of explanations are of the same type—they all refer to mechanisms. As explained in Sects. 2.3 and 2.4, the different variants of mechanistic explanation differ with regard to the nature of their explananda (for an overview see Fig. 2.4) as well as the nature of the relation between the mechanisms and the phenomenon—on which I will say more on below.

Output mechanistic explanations are causal explanations and what is standardly called “etiological mechanistic explanations”. The explanandum refers to an event that is an effect of a mechanism. The relation between mechanism and phenomenon/explanandum and explanans is causation. The relevant question is “How does this effect come about?” For example, when scientists explain neurotransmitter release, they refer to the mechanism that causes neurotransmitter release.

Input-output mechanistic explanation is a type of etiological explanation as well. The phenomenon consists of a set of causes/inputs and effects/outputs that are connected by the mechanism. The relation between the mechanism and the explanandum phenomenon is that of causation. Indeed, input-output mechanistic explanations

are common in the life sciences. Explanatory requests asking for input-output mechanistic explanations are, for example, “How does the release of neurotransmitters at the axon terminal change depending on changing inputs to the pre-synaptic neuron?” or “How is neurotransmitter release generated if an action potential reaches the axon terminal?”.

In *filler mechanistic explanations*, the explanandum is characterized in terms of a functional role. This functional role term functions as a *black box* or a *filler term*. The new mechanists talk of “black boxes” that scientists refer to by using “filler terms” that are filled as researchers gain more and more knowledge about the underlying mechanisms (i.e., with details about the mechanism) (Craver, 2007b; Piccinini & Craver, 2011). Scientists often start from “some-process-we-know-not-what” (Craver, 2007b, p. 114) that is specified by an input-output relation in order to find out which mechanism connects the inputs and the outputs. For example, protein synthesis may be characterized as whatever process that starts with DNA molecules being separated into two strands by the RNA polymerase and ends with there being new proteins; then the mechanism that realizes this input-output relation is found, which is then identified with protein synthesis. In other words, “protein synthesis” is a filler term denoting a black box in a mechanistic model involving the existence of newly produced proteins. This black box is filled by gathering more and more knowledge about the entities and activities that are engaged in protein synthesis (the mechanism that leads to there being new proteins). Protein synthesis, then, is identified with the mechanism. The relevant research question here is “What are the components of the process-we-don’t-know-yet?” The process-we-don’t-know-yet is referred to by a filler-term. Explanatory requests that ask for filler mechanistic explanations are “How does *filler term* work?” or “How are the outputs generated given the inputs?”. Given that filler terms are specified in terms of an input-output relation, the two questions are basically the same.

In *dimensioned mechanistic explanations*, derived from the behaving entity view, the explanandum is the behavior of an object that contains the mechanism. Here, the relevant question is “How does this object do THIS?” This kind of mechanistic explanation comes into play when we explain, for example, how muscles contract, how mice navigate the Morris Water maze, or how neurons fire. They inherit their name “*dimensioned* mechanistic explanation” from so-called *dimensioned accounts of realization*. Dimensioned views of realization are standardly contrasted with flat views of realization. The difference between these accounts is whether they take realization to be a relation between properties of *one and the same* individual (“flat”) or of *two different individuals* (“dimensioned”) (Endicott, 2011; Gillett, 2010). Since the explananda and the explanantia of dimensioned mechanistic explanation refer to (activities and properties of) different individuals, whereas filler mechanistic explanations identify the explanandum and the explanans, I call this version of mechanistic explanation “dimensioned”.

2.6 Constitutive Mechanistic Explanation: Different Ontological Implications

The new mechanistic approach does not only deliver an analysis of scientific explanation. In the context of the mechanistic framework, ontological questions are discussed that are directly related to the nature of mechanistic explanation. I want to highlight three:

1. Are the explanations of higher-level sciences reducible to the explanations of lower-level sciences?
2. What are levels?
3. Can there be causal relations between levels?

Constitutive mechanistic explanation is taken to be the explanatory type that is crucial for addressing these questions because mechanistic constitution is taken to be the relation that holds between levels. To answer the questions, thus, one needs a proper understanding of what constitutive explanation is. Since, as I have shown in the preceding sections, there are two variants of constitutive mechanistic explanation, the discussions of the three questions must pay attention to the different types of mechanistic explanation. In what follows, I highlight some implications of the different explanation types regarding questions (1), (2), and (3).

1. Constitutive mechanistic explanation and reduction

There are various different ways of asking the reduction-question, depending on (a) which relata one is interested in (e.g., disciplines, theories, explanations, entities), (b) what one takes the contrast to be (e.g., reduction vs. autonomy, reduction vs. integration), and (c) what one takes the relevant relation between the relata to be that justifies or blocks reduction (e.g., the impossibility of Nagel-reduction, multiple realization, impossibility of decomposition). Here, I will only briefly outline the different implications of filler and dimensioned mechanistic explanation regarding the question of whether phenomena are reducible to mechanisms.

There are two variants of reduction that need to be kept apart here: *ontological reduction* and *epistemic reduction*. The former concerns the question of whether the phenomenon de facto is nothing but the mechanism or not. The latter concerns the question of whether all knowledge about the phenomenon is nothing but or can be derived from knowledge about the mechanism. Regarding ontological reduction, the implications of filler mechanistic explanation and dimensioned mechanistic explanation are already clear. Since filler mechanistic explanation implies that the phenomenon is identical to the mechanism (see Sect. 2.3), and identity is usually taken to be sufficient for ontological reduction, filler mechanistic explanation implies that the phenomenon is ontologically reducible to the mechanism. In contrast to that, according to dimensioned constitutive explanation, phenomena are not ontologically reducible to their underlying mechanisms. Mechanisms, according to this view, are *proper* parts of phenomena. For example, the navigation mechanism

involves only some but not all parts of the navigating mouse. Furthermore, the properties of the phenomenon are “qualitatively distinct” from the properties of the mechanism (see Sect. 2.4). Navigation mechanisms by themselves can’t navigate. Only the system as a whole of which the navigation mechanisms is a part of can do so.

Regarding epistemic reduction, Bechtel—focusing on disciplines rather than phenomena and mechanisms—highlights that higher levels provide information that the corresponding lower levels do not contain: namely *organizational* and *contextual* information (Bechtel, 2007, pp. 182–183). This, however, cannot be true for phenomena that are explained by filler mechanistic explanations. As Fazekas and Kertész argue, “[o]nce the required identity statements are in place one is able to infer to higher level processes from lower level knowledge” (2011, pp. 380–381) (see also Krickel (2018a, p. 117)). All there is to know about the phenomenon, based on filler mechanistic explanation, is the respective causal profile. Since the causal profile just is the causal profile of the underlying mechanism, all knowledge about the phenomenon is included in the knowledge about the mechanism. The situation is different for dimensioned mechanistic explanation. Here, the phenomenon is much richer than the mechanism. The phenomenon is a behavior of a system that contains more than just the mechanism. The same holds for the behavior of the system that is to be explained. For example, the mouse’s navigation behavior may de facto involve, say, the mouse being exhausted at a certain time. This is nothing we can derive from knowing the navigation mechanisms. Furthermore, the mouse’s navigation behavior occurs in a certain environment. Again, knowledge about the environment cannot be derived from the knowledge about the mechanism. It may even be that the navigating mechanism does exactly the same in two different environments.

In a nutshell: in filler mechanistic explanation, the phenomenon is ontologically and epistemically reducible to the mechanism. In dimensioned mechanistic explanations, the phenomenon is neither ontologically, nor epistemically reducible to the mechanism.

2. Constitutive mechanistic explanation and levels of mechanisms

The two different variants of constitutive mechanistic explanations have different implications regarding the interpretations of levels of mechanisms. According to the new mechanists, x is at a lower mechanistic level than y if and only if y is a component in the mechanism for x . In other words, the components of the mechanism for a phenomenon P are at a lower mechanistic level than P . If the phenomenon just is the mechanism, it follows that the entities and activities that make up a mechanism are at a lower level than the mechanism itself. That is, the relata of the level-relation, according to filler mechanistic explanation, are the mechanism and its components. For example, according to this view, the hippocampus generating spatial maps is at a lower level than the spatial navigation mechanism. And the opening ion-channel is at a lower level than the mechanism that is responsible for the propagation of the

action potential along the axon.⁴ Based on filler mechanistic explanation, mechanistic levels arise due to *organization*. A bunch of acting entities alone does not give rise to a new mechanistic level. Only if they are put into the right kind of temporal, spatial, and causal organization, they form a mechanism, and thereby create a new level.

Dimensioned mechanistic explanation provides a different picture. Here, the relata are the behaving system and the mechanism's components. For example, according to this view, the hippocampus generating spatial maps is at a lower level than the navigating mouse. The ion channel opening is at a lower mechanistic level than the firing neuron. According to this picture, it is the containment-relation plus organization that gives rise to new level of mechanisms. A bunch of acting entities put into a certain organization *and* put into the context of a larger system (e.g., an organism) creates a new mechanistic level—because the larger system will show a new behavior that it would not be able to perform without the acting entities in that specific organization and that the acting entities in that organization could not do by themselves.

3. Constitutive mechanistic explanation and interlevel causation

The different pictures of mechanistic levels suggested by filler mechanistic explanation and dimensioned mechanistic explanation have different implications for the possibility and nature of interlevel causation. According to filler mechanistic explanation, interlevel causation would require that the mechanism as a whole would causally interact with its components. According to dimensioned mechanistic explanation, interlevel causation would hold between the behaving larger system and the mechanistic components.

At a first glance, interlevel causation between mechanistic levels is excluded for almost trivial reasons. A standard assumption about causation is that it relates events that occur at different times—i.e., causes are standardly assumed to precede their effects. On both views, however, the relata of mechanistic levels are wholes (mechanisms, larger behaving systems) and their parts (mechanistic components). If, however, one thinks of mechanisms and behaving systems as temporally extended things that have different temporal phases (Krickel, 2017), then this worry can be solved. One could think of interlevel causation as holding between the mechanism or the behaving system at t_m and a mechanistic component at t_n (where $m \neq n$). On this assumption, interlevel causation in the filler mechanistic picture would, however, be identical to same-level causation. The reason is that each temporal phase of the

⁴Peter Fazekas (2022), however, argues that the mechanism and its components are not at distinct levels and that the notion of a mechanistic level, thus, needs to be rejected. According to Fazekas, the description of the mechanism as a whole and the description of the components just “provide different levels of description of the very same phenomenon” (Fazekas, 2022, 2310). However, Fazekas seems to take the relata of the level relation to be mechanisms, on the one hand, and the organized components, on the other. However, this really seems to amount to double counting: the mechanism just is the collection of its components in a certain organization (this is what MC expresses, see Sect. 2.2). This, however, is compatible with the mechanistic view that each component is at a lower level than the mechanism of which it is a part.

mechanism just is the interaction of the mechanism's components at the given time. For example, whatever the navigation mechanism does at t_i —it just is what the components that make up the navigation mechanism at t_i are doing at t_i . Thus, the claim that the navigation mechanism at t_i causes, say, the hippocampus's activity at t_j would be simply translated to the claim that the mechanistic components at t_i cause the hippocampus's activity. The different components of the navigation mechanism, however, are not at different levels of mechanisms. Thus, on the filler mechanistic picture, there is no interlevel causation in mechanism.

Again, dimensioned mechanistic explanation provides a different picture. Remember that we described causation between different level of mechanisms as a causal relation between a temporal phase of the higher-level at t_m and temporal phase of the lower-level at t_n . On this picture, this would be, for example, the navigation behavior of the mouse at t_m and the hippocampus's activity at t_n . The temporal phases of the mouse's navigation behavior are not just the interactions of the mechanistic components. These temporal phases are, for example, the mouse's turning left, the mouse's stopping, or the mouse's running faster. It is in line with the overall picture to say that the mouse's turning left at t_m is at a higher level than the hippocampus activity at t_n (for an argument in that direction see Krickel (2017)). Furthermore, *prima facie*, it makes sense to say that the mouse's turning left at t_m is a cause of the hippocampus's activity at t_n . Whether this is indeed true depends on what exactly one takes causation to be and, of course, on empirical facts.

In a nutshell: there cannot be interlevel causation between the phenomena and the mechanistic components as understood in filler mechanistic explanations. Dimensioned mechanistic explanation provides a picture of phenomena and mechanistic components that in principle could be causally related.

2.7 Conclusion

I have argued that there are indeed four different types of mechanistic explanation, two of which could be summarized under the label "etiological mechanistic explanation" and the other two as "constitutive mechanistic explanation". This insight follows from taking a closer look at the different assumptions that have been made in the mechanistic literature on constitutive mechanistic explanation. Based on Krickel (2018a) and Kaiser and Krickel (2017), I have shown that there are three different types of explanation that might (mistakenly) all be subsumed under the label of constitutive mechanistic explanation: input-output mechanistic explanation, filler mechanistic explanation, and dimensioned mechanistic explanation. While all of these types of explanation can indeed be found in the life sciences, only the latter two exemplify the features that are standardly attributed to constitutive mechanistic explanation. Input-output mechanistic explanation, indeed, is a variant of etiological mechanistic explanation. Furthermore, I have shown that the two acceptable views of constitutive mechanistic explanation have different implications regarding reduction, levels of mechanism, and interlevel causation.

References

- Baetu, T. M. (2012). Filling in the mechanistic details: Two-variable experiments as tests for constitutive relevance. *European Journal for Philosophy of Science*, 2(3), 337–353. <https://doi.org/10.1007/s13194-011-0045-3>
- Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*, 67(3), 731–756. <https://doi.org/10.1093/bjps/axv003>
- Bechtel, W. (2007). Reducing psychology while maintaining its autonomy via mechanistic explanations. In M. Schouten & H. Looren de Jong (Eds.), *The matter of the mind: Philosophical essays on psychology, neuroscience and reduction* (pp. 172–198). Basil Blackwell.
- Bechtel, W. (2008). Mental mechanisms. In *Philosophical perspectives on cognitive neuroscience*. Routledge.
- Bechtel, W. (2017). Explicating top-down causation using networks and dynamics. *Philosophy of Science*, 84(2), 253–274. <https://doi.org/10.1086/690718>
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421–441. <https://doi.org/10.1016/j.shpsc.2005.03.010>
- Casini, L., Illari, P. M., Russo, F., & Williamson, J. (2011). Models for prediction, explanation and control: Recursive Bayesian networks. *Theoria-Revista De Teoria Historia Y Fundamentos De La Ciencia*, 26(1), 5–33. <https://doi.org/10.1387/theoria.784>
- Couch, M. B. (2011). Mechanisms and constitutive relevance. *Synthese*, 183(3), 375–388. <https://doi.org/10.1007/s11229-011-9882-z>
- Craver, C. F. (2007a). Constitutive explanatory relevance. *Journal of Philosophical Research*, 32(Section II), 3–20. <https://doi.org/10.5840/jpr20073241>
- Craver, C. F. (2007b). *Explaining the brain: Mechanisms and the mosaic Unity of neuroscience*. Oxford University Press.
- Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 199(3–4), 8807–8828. <https://doi.org/10.1007/s11229-021-03183-8>
- Endicott, R. P. (2011). Flat versus dimensioned. *Journal of Philosophical Research*, 36, 191–208. https://doi.org/10.5840/jpr_2011_13
- Fagan, M. B. (2012). The joint account of mechanistic explanation. *Philosophy of Science*, 79(4), 448–472. <https://doi.org/10.1086/668006>
- Fazekas, P. (2022). Flat mechanisms: A reductionist approach to levels in mechanistic explanations. *Philosophical Studies*, 179(7), 2303–2321. <https://doi.org/10.1007/s11098-021-01764-4>
- Fazekas, P., & Kertész, G. (2011). Causation at different levels: Tracking the commitments of mechanistic explanations. *Biology and Philosophy*, 26(3), 365–383. <https://doi.org/10.1007/s10539-011-9247-5>
- Gillett, C. (2010). Moving beyond the subset model of realization: The problem of qualitative distinctness in the metaphysics of science. *Synthese*, 177(2), 165–192. <https://doi.org/10.1007/s11229-010-9840-1>
- Gillett, C. (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds and Machines*, 23(3), 309–337. <https://doi.org/10.1007/s11023-013-9311-9>
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44(1), 49–71. <https://doi.org/10.1007/BF00172853>
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342–S353. <https://doi.org/10.1086/341857>
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press. <https://doi.org/10.1093/oso/9780198779711.001.0001>
- Harbecke, J. (2010). Mechanistic constitution in neurobiological explanations. *International Studies in the Philosophy of Science*, 24(3), 267–285. <https://doi.org/10.1080/02698595.2010.522409>

- Harinen, T. (2018). Mutual manipulability and causal inbetweenness. *Synthese*, 195(1), 35–54. <https://doi.org/10.1007/s11229-014-0564-5>
- Illari, P. M., & Williamson, J. (2011). Mechanisms are real and local. In *Causality in the sciences* (pp. 818–844). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199574131.003.0038>
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2(1), 119–135. <https://doi.org/10.1007/s13194-011-0038-2>
- Irvine, E. (2013). *Consciousness as a scientific concept*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-5173-6>
- Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science*, 68(3), 745–779. <https://doi.org/10.1093/bjps/axv058>
- Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology and Philosophy*, 27(4), 545–570. <https://doi.org/10.1007/s10539-012-9308-4>
- Kästner, L. (2017). *Philosophy of cognitive neuroscience, causal explanations, mechanisms and experimental manipulations*. De Gruyter. <https://doi.org/10.1515/9783110530940>
- Kirchhoff, M. D. (2017). From mutual manipulation to cognitive extension: Challenges and implications. *Phenomenology and the Cognitive Sciences*, 16(5), 863–878. <https://doi.org/10.1007/s11097-016-9483-x>
- Kistler, M. (2009). Mechanisms and downward causation. *Philosophical Psychology*, 22(5), 595–609. <https://doi.org/10.1080/09515080903238914>
- Krickel, B. (2017). Making sense of interlevel causation in mechanisms from a metaphysical perspective. *Journal for General Philosophy of Science*, 48(3), 453–468. <https://doi.org/10.1007/s10838-017-9373-0>
- Krickel, B. (2018a). *The mechanical world* (Vol. 13). Springer International Publishing. <https://doi.org/10.1007/978-3-030-03629-4>
- Krickel, B. (2018b). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A*, 68, 58–67. <https://doi.org/10.1016/j.shpsa.2018.01.003>
- Kuorikoski, J. (2012). Mechanisms, modularity and constitutive explanation. *Erkenntnis*, 77(3), 361–380. <https://doi.org/10.1007/s10670-012-9389-0>
- Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science*, 63(2), 399–427. <https://doi.org/10.1093/bjps/axr036>
- Levin, J. (2023). Functionalism. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Summer 2). Metaphysics Research Lab, Stanford University.
- Levy, A. (2009). Carl F. Craver, explaining what? Review of explaining the brain: mechanisms and the mosaic unity of neuroscience. *Biology & Philosophy*, 24(1), 137–145. <https://doi.org/10.1007/s10539-008-9123-0>
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17), 556–567. <https://doi.org/10.2307/2025310>
- Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183(3), 1–58. <https://doi.org/10.1007/s11229-011-9898-4>
- Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 192(11), 3731–3755. <https://doi.org/10.1007/s11229-015-0718-0>
- Salmon, W. C. (1984a). Scientific explanation: Three basic conceptions. *PSA: Proceedings of the biennial meeting of the philosophy of science association, 1984*, 293–305.
- Salmon, W. C. (1984b). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Soom, P. (2012). Mechanisms, determination and the metaphysics of neuroscience. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(3), 655–664. <https://doi.org/10.1016/j.shpsc.2012.06.001>

- van Eck, D., & Looren de Jong, H. (2016). Mechanistic explanation, cognitive systems demarcation, and extended cognition. *Studies in History and Philosophy of Science Part A*, 59(Supplement C), 11–21. <https://doi.org/10.1016/j.shpsa.2016.05.002>
- Weinberger, N. (2019). Mechanisms without mechanistic explanation. *Synthese*, 196(6), 2323–2340. <https://doi.org/10.1007/s11229-017-1538-1>
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
- Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research*, 91(2), 303–347. <https://doi.org/10.1111/phpr.12095>
- Zednik, C. (2015). *Heuristics, descriptions, and the scope of mechanistic explanation* (pp. 295–318). Springer. https://doi.org/10.1007/978-94-017-9822-8_13

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 3

The Metabolic Theory of Ecology as a Mechanistic Approach



Gonçalo Martins

Abstract Philosophy of science has recently given a great deal of attention to the concept of mechanism. However, unlike the biological mechanisms identified in other fields of the life sciences, ecological mechanisms have not been exhaustively examined. The aim of this chapter is to critically analyze the Metabolic Theory of Ecology. This theory is supposed to provide a unification of population, community and ecosystem approaches rooted in the ecophysiology of individual organisms. In this context, metabolism plays a fundamental role as the unifying concept between levels. According to its authors, this is a mechanistic approach to ecology involving decomposability into parts that structure the different levels of ecological organization and into mechanisms that can be characterized by identifying a phenomenon, parts, causing, and organization. I shall first argue that its mechanistic nature needs clarification. I shall then suggest that the theory can explain some phenomena at various levels of ecological organization and can describe some patterns or tendencies in nature, although it is not able to completely elucidate their mechanistic basis, i.e., to explain the mechanisms that produce these patterns.

Keywords Mechanism · Metabolism · Explanation · Allometry · Scaling

3.1 Introduction

The concept of mechanism has recently received a great deal of attention in the philosophy of science. The life sciences offer philosophers of science a variety of examples to challenge the more traditional deductive-nomological model of explanation, in which explanation is provided by derivations from laws. Different areas of biology indicate that scientific enquiry is driven by a search for mechanisms and

G. Martins (✉)

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências,
Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
e-mail: fc33113@alunos.ciencias.ulisboa.pt

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of
the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_3

29

that explanation is a matter of characterizing them in specificity (Bechtel & Richardson, 2010; Illari & Williamson, 2012; Craver & Darden, 2013; Raerinne, 2013; Pâslaru, 2018). Many biologists and ecologists effectively use a mechanistic research perspective, looking for mechanisms conceptualized as entities generating a phenomenon to be explained. The most basic concept of mechanism – as a start-to-finish sequence of qualitatively characterized operations performed by component parts – was provided by Machamer et al. (2000). This linear characterization is sufficient to single out some important epistemic practices by means of which scientists decompose mechanisms *structurally* into their *parts* and *functionally* into their *operations*. Glennan (1996) has also developed a mechanistic account, which, initially, retained the centrality of laws in order to explain the interaction of parts. More recently, Glennan (2002) has replaced the language of laws with that of “invariant change-relating generalizations”, an approach that I will also develop further in Sect. 3.4, when the putative mechanistic nature of the Metabolic Theory of Ecology (henceforth, MTE) is analyzed. Bechtel characterizes a mechanism as “a structure performing an action in virtue of its component parts, component operations, and their organization”, adding that “the orchestration of the mechanism is responsible for one or more phenomena” (Bechtel, 2006: 26). Machamer et al. (2000) rejected Glennan’s emphasis on interactions, something that is also relevant for Bechtel and Richardson (2010), and emphasized the dualism of entities and activities. Bechtel and Richardson (2010: 24) had already made implicit this dualism in their discussion of decomposition and localization within the mechanistic account, but they rejected the underlying linearity in Machamer et al. (2000) perspective and, using Glennan’s language of properties, have developed a more dynamical account of mechanistic explanation, assuming patterns of change over time in the properties of the parts and operations. More recently, Glennan et al. (2021: 145) have argued for a *Minimal Mechanism Thesis*, according to which “a mechanism for a phenomenon consists of entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon”. According to these authors, this thesis identifies some points of consensus about what mechanisms are because they are identified and individuated by the phenomena they explain, by the entities or parts they are made of, by what their activities or interactions do, and by the organization in which they are structured.

The change in focus from characterizing explanation in terms of derivation from laws to understanding the role of mechanisms in generating phenomena provides a very different perspective. Let me briefly explain in what sense mechanistic explanation differs from the outdated – at least for many mechanists - deductive-nomological model of explanation (for a more historical overview, see Nicholson, 2011; Illari & Williamson, 2012).

Firstly, the crucial component of a mechanistic account is not the formulation of the relevant law; it is, instead, the determination of the parts of the mechanism, the operations they perform, and how they are organized (Bechtel, 2011; Craver & Tabery, 2019; Glennan et al., 2021). Secondly, although these parts and operations can be described linguistically, it is often more productive to represent them in diagrams (Bechtel, 2011). Thirdly, the demonstration that the mechanism can produce

the phenomenon does not rely on logical derivations but, rather, on mental simulations of the mechanism in operation, later ascertained by empirical research. Fourthly, mechanistic explanations are inherently reductionist insofar as they require specifying the parts of a mechanism and the operations they perform (Craver & Tabery, 2019; Pâslaru, 2018). However, they also require consideration of the organization of the whole mechanism and its relation to conditions in its environment, since it is only when appropriately situated that a mechanism will produce the phenomenon of interest (Bechtel & Abrahamsen, 2005).

Mechanisms are *decomposable* in the sense that the behavior of a system as a whole can be broken down into organized interactions among the parts. There are numerous characterizations of mechanisms in the literature, and a “consensus concept” might be adopted: “A mechanism for a phenomenon consists of entities (or parts) and activities (or operations) organized in such a way that they are responsible for the phenomenon”.¹ All characterizations contain four basic features: a phenomenon, parts, causing, and organization. The *phenomenon* is the behavior of the mechanism as a whole, being all mechanisms the mechanisms for some phenomenon. The boundaries of a mechanism are fixed by reference to the phenomenon that the mechanism explains. The parts in a mechanism are component parts in virtue of being relevant to explaining the phenomenon.² There has been a great struggle to find a concise way to express the idea of what a *part* is, a crucial concept required to define the components of a mechanism.³ Mechanists have disagreed with one another about how to understand the concept of mechanistic *cause* (Craver & Tabery, 2019). New mechanists have been at pains to liberate the relevant causal notion from any overly austere view that restricts causation to only a small class of phenomena, generally associated with physics, such as collisions, attractions or repulsions. Another difficulty has been to distance themselves from the regularist conception of causation, in the Humean sense, common among the logical

¹This characterization is labeled a “consensus concept” in the papers of Nicholson (2011) and Illari and Williamson (2012), and by the *Stanford Encyclopedia of Philosophy* entry (Craver & Tabery, 2019).

²New mechanists speak variously of the mechanism as producing, underlying, or maintaining the phenomenon (Craver & Tabery, 2019). The language of production is probably best applied to mechanisms conceived as a causal process terminating in some end product (e.g. protein biosynthesis). In contrast, for physiological mechanisms, it is more appropriate to say that the mechanism underlies the phenomenon. For ecology, the preferred idea is that a mechanism might *maintain* a phenomenon, in a homeostatic sense. In this case, the phenomenon is a state of affairs, or a range of states of affairs, that is held in place by the mechanism. An area of active discussion is whether the relationship between the mechanism and the phenomenon must be regular. Machamer et al. (2000: 3) stipulate that mechanisms are regular, in that they work “always or for the most part in the same ways under the same conditions”, a position that is of prominent use in ecology.

³Formal mereologies are difficult to apply to the material parts of mechanisms in the life sciences. The axioms of mereology, such as reflexivity (i.e., everything is a part of itself) and unrestricted composition (i.e., any two things form a whole) do not apply in standard life sciences’ uses of the “part” concept (Craver & Tabery, 2019). Glennan’s proposal (1996: 53), for example, is that: “The parts of mechanisms must have a kind of robustness and reality apart from their place within that mechanism. It should in principle be possible to take the part out of the mechanism and consider its properties in another context”.

empiricists.⁴ The characteristic *organization* of mechanisms is also the subject of considerable discussion (Wimsatt, 1997; Machamer et al., 2000; Bechtel, 2011; Craver & Tabery, 2019). It is relevant to contrast mechanistic organization with *aggregation*, a distinction that mechanists have used to articulate how the parts of a mechanism are organized together to form a whole. This distinction is crucial in the analysis of some reductionist mechanistic approaches to ecology. Aggregate properties are properties of wholes that are simple sums of the properties of their parts. In aggregates, the parts can be rearranged and intersubstituted without changing the property or behavior of the whole; the whole can be taken apart and put back together without disrupting its behavior, and the properties of the whole change linearly with the addition and removal of parts. Organization can be conceived as non-aggregativity, allowing a mechanistic form of emergence (Wimsatt, 1997). Mechanists have also detailed several kinds of organization characteristic of mechanisms (Craver & Tabery, 2019), in particular spatial organization (including location, shape and orientation) and temporal organization (including order, rate and duration of component operations). Other important features of organization are modularity, a property characterizing the relative functional independence of some parts, i.e., meaning that it should be physically possible to intervene on a putative cause variable of a mechanism without disrupting the functional relationships among the other variables; jointness, a property related to modularity that characterizes the interdependent relationship between parts of a mechanism in the sense that components in a mechanism often form a more complex unit by virtue of the individual properties uniting them; and, finally, mechanists emphasize the hierarchical organization of mechanisms and the multilevel structure approaches in the special sciences, such as ecology, demanding an analysis of mechanistic relations across levels of organization (Craver & Tabery, 2019).

Concerning the issue of explanation, while in the deductive-nomological model explanations are considered arguments showing that the event to be explained is to be expected on the basis of the relevant laws of nature and antecedent and boundary

⁴Four ways of discussing the concept of cause have been prominent in this debate (Bechtel, 2011; Craver & Tabery, 2019). The first way defends a transmission account in which causation involves the transmission and propagation of marks or conserved quantities. This view has been unpopular in the life sciences because, within this domain, causal claims usually do not involve explicit reference to conserved quantities. The second account sees causation as derivative from the concept of mechanism, in the sense that causal claims are claims about the existence of a mechanism. The truth-maker for a causal claim at one level of organization is a mechanism at a lower level, that is, mechanisms are the hidden connection Hume sought between cause and effect. The third account embraces the view that causation should be understood in terms of productive activities. This account has been criticized because it fails to say what activities are and to account for the relationship of causal and explanatory relevance. The last account has the central commitment that models of mechanisms describe variables that make a difference to the values of other variables in the model and to the occurrence of a phenomenon. Difference-making in this manipulationist sense is understood as a relationship between variables in which interventions on cause variables can be used to change the value of effect variables. This is the replacement that Glennan (2002) has made, changing the language of laws with that of “invariant change-relating generalizations”, an approach that I shall explore in more detail in Sect. 3.4, following the argument of Raerinne (2011).

conditions, mechanists, in contrast, insist that explanation is a matter of elucidating the causal structures that produce, underlie, or maintain the phenomenon of interest (Illari & Williamson, 2012; Nicholson, 2011; Craver & Tabery, 2019). Thus, the philosophical problem is largely about characterizing or describing the worldly or ontic structures to which explanatory models must refer to if they are to count as genuinely explanatory. The phenomenon must be situated within the causal structure of the world, and the explanation is an account of how the phenomenon is produced by entities and their properties. However, it is important to emphasize that, even if mechanistic explanations were ubiquitous across empirical sciences, this fact does not entail necessarily that all scientific explanations must be mechanistic, even though some critics claimed that New Mechanists are committed to such a radical position (Glennan et al., 2021).

Most mechanists recognize two main aspects of mechanistic explanation (Craver & Tabery, 2019): the etiological, which reveal the causal history of the *explanandum* phenomenon; and the constitutive, which explain a phenomenon by describing the mechanism underlying it. With increased attention to the latter, mechanists realized the need for an account of constitutive relevance, a principal need for sorting relevant from irrelevant factors within a mechanism. One central research problem is therefore to identify which of these entities, activities and organizational features contribute to the phenomenon and which do not. In a sense, the challenge is to define the boundaries of a mechanism, i.e., of saying what lies within and outside the mechanism.

It is relevant that in much of the literature on mechanisms, these are contrasted explicitly with laws of nature (Machamer et al., 2000; Craver & Darden, 2013; Craver & Tabery, 2019). This contrast grew out of an emerging consensus in philosophy of science that there are few, or perhaps no, laws in the life sciences (Lawton, 1996; Colyvan, 2003). The empirical generalizations found in biology and ecology tend to be hedged by *ceteribus paribus* clauses; whether they hold or not depends on background conditions that might not hold and on conditions internal to the mechanism that might fail to occur. In short, these generalizations are mechanistically explicable, and whatever necessity they might possess derives from a mechanism. Thus, mechanisms seem to play the role of laws in the life sciences, because one seeks mechanisms to explain, predict and control phenomena in nature even if mechanisms lack many of the characteristics that define laws in the logical empiricist framework, such as universality, inviolable necessity or unrestricted scope.

Research on mechanisms has also helped to clarify the idea of levels of organization and its relation to other forms of organization and non-mechanistic forms of emergence. Many mechanists emphasize that biological and ecological systems are hierarchically organized into near-decomposable⁵ structures (see Craver & Tabery,

⁵They are nearly decomposable because, sometimes, in the life sciences, it is extremely difficult to identify all the components of a mechanism. Wimsatt, following the work of Herbert Simon, offered a treatment of the levels of nature in which, accordingly, hierarchical systems are organized around interrelated subsystems, with discriminate capacities (Bechtel & Richardson, 2010). The idea is that not all systems exhibit interesting forms of hierarchy, but many are hierarchically organized, and that "(...) among those that do exhibit some degree of hierarchical order, the key feature

2019, for a description of this perspective): mechanisms within mechanisms. Bechtel and Richardson (2010) argue that two essential claims concerning hierarchical organization can be distinguished: one states that nature is organized in terms of wholes and parts, mereologically; the other asserts that organization sometimes is hierarchical without being decomposable, but when it is decomposable, or nearly so, there is a very natural ranking of types of systems. Thus, when we have a case of strict decomposition, one can project the component part behavior to the systemic behavior; if we have a case of near decomposability, the component part behavior can be approximately projected to the systemic behavior; finally, even in the case of non-decomposition – in which the behavior of one part depends on the behavior of the other parts, as in ecology – assuming the decomposability of a system “can be highly informative” (Bechtel & Richardson, 2010: xxix). These authors defend that *decomposition* and *localization* are powerful heuristics for discovering mechanisms and for articulating their *structure*; however, in the case of the life sciences, they emphasized *functional* decomposition and localization, assuming that the system responsible for some phenomenon is hierarchical and decomposable – that is, it results from different parts within the mechanism performing their activities. Accordingly, researchers should aim to decompose the phenomenon into the component operations that produce it and localize them within the parts of the mechanism. This is the perspective adopted by the advocates of the MTE, when they decompose the ecosystem’s metabolism in the metabolic processes of populations and individual organisms, as I shall illustrate in Sect. 3.2. However, Bechtel and Richardson (2010) recognize that decomposition can be challenging for those researching natural systems because mechanisms usually do not reveal their parts, while the component operations can be even harder to differentiate. These authors also admit that the notion of localization can be criticized, with the assumption that specific activities can be localized in discrete parts of a mechanism. Nevertheless, Bechtel and Richardson (2010) defend a different construal of localization, which is neither direct nor simple, and whose goal is not to find where an activity takes place but to acquire information about the part involved. The conception of localization of Bechtel and Richardson (2010) is rather different from that assumed by their critics. First, the authors recognize that, although sometimes researchers begin by assuming that the activity of a mechanism is due to one component within it (*direct* or *simple* localization), this is only a preliminary step in research: once the component is identified and its behavior explored, it turns out not to generate the phenomenon on its own but to perform an operation that, together with the other operations performed by other components, generates the phenomenon. As research proceeds, it is not the whole phenomenon that is localized in a part of the system but, rather,

concerns the relative strength of interaction among as opposed to within subsystems. The clear and straightforward thought is that as interaction among subsystems increases in importance, the significance of interaction within subsystems decreases, and vice versa. So, in relatively simple hierarchies, there will be a relatively high strength of interaction within subsystems as compared with the interaction among subsystems. These are systems that are at least “nearly decomposable” (Bechtel & Richardson, 2010: xxix).

individual operations, each of which contributes in some way to the phenomenon of interest. Second, although the authors think that the word localization suggests a single discrete spatial location, that is not necessary and is often not correct, because the functional component may be distributed in space, with the intervening space containing parts performing other operations or even entities that are not part of the mechanism responsible for the given phenomenon. A further criticism of localization is that identifying the part performing an operation is of no intrinsic interest. This points to the third, and probably most important, difference in the construal of localization by Bechtel and Richardson (2010): they treat localization not as an end of inquiry but as a heuristic. The goal of localizing is not only to find out where something occurs but to acquire information about the part that is engaged in that operation, which can inform further research. Thus, when direct or simple localization proves inadequate for understanding a given phenomenon, and the phenomenon is functionally decomposed into multiple operations localized in different parts, the issue of how these parts and operations are organized becomes important. In these authors' perspective, researchers begin with simpler conceptions and hypothesize that a mechanism comprises component parts whose operations are performed sequentially. For the authors the simpler hypothesis may not be the best, but it is after all the simplest, making it easier to assess it. This is what they characterize as *indirect* or *complex* localization (a concept already envisaged in Machamer et al.'s (2000) linear characterization of a mechanism). Even when one knows that a simple model is not defensible, the attempt to understand mechanisms in terms of a linear execution can be very productive.

This view leads to the conclusion that evolved structures, such as biological and ecological ones, are more likely to be nearly decomposable into hierarchically organized, more or less stable structures and sub-structures. An important objection has been raised against this perspective (see Bechtel & Richardson, 2010; Craver & Tabery, 2019, for an overview of the discussion), – first by the vitalists and organic holists in the nineteenth and early twentieth century and, more recently, by certain dynamicists – stating that it is misleading because evolution does not construct natural systems from scratch, piece by piece. However, there have been some attempts to reconstruct this argument and tackle this kind of criticism, as a way of showing that evolved systems are more likely to be modular: systems made of independently manipulable parts that can quarantine the effects of changes to specific parts, giving them flexibility to make local changes without causing great side-effects. It is important to add that scientists can only describe and explain mechanisms through the construction of models, which are representations of mechanisms. Such representations are inevitably partial, abstract, idealized and plural (Glennan et al., 2021). Accordingly, the crucial point is that mechanist philosophers, whether they think that explanations are epistemic entities, such as models or representations, or that explanations possess a preponderant ontic component, they should grant that the models are required to provide good mechanistic explanations. This discussion will be relevant for the analysis of the virtues and limitations of MTE, in Sect. 3.3.

The near decomposability of mechanisms is directly related to the idea that mechanisms span multiple levels of organization. The behavior of the whole is

explained in terms of the activities and interactions among the component parts. These activities and interactions are themselves sustained by underlying activities and interactions among component parts, and so on. Levels of mechanisms can be defined in terms of a relationship between the behavior exhibited by a system and the activity of some component part of that system (Craver & Bechtel, 2007). On this account, the activity of a component is at a lower level of mechanistic organization than the behavior of the system if and only if the component is a part of the system, and its activity is part of the system's behavior. In short, to say that something is at a lower mechanistic level than the mechanism as a whole is to say that it is a working part of the mechanism.

For some mechanists, one implication of this view of levels, combined with certain familiar assumptions about causal relations, is that there can be no causal relationships between items at different levels of mechanisms (see Craver & Bechtel, 2007; Craver & Tabery, 2019, for a description of this posture). This is a position in contrast with some form of holism, which necessarily does not limit the analysis to the constitutive parts of – or their relations within – a specific level of organization. According to holism, both the higher levels (“downward causation”) and the lower ones (“upward causation”) might participate in determining the properties of specific levels. According to non-holist mechanists, claims about inter-level causation concerning mechanisms are expressed as hybrid claims combining, on the one hand, constitutive claims about the relationship between the behavior of the mechanism as a whole and the activities of its parts and, on the other hand, causal claims concerning relationships between things not related as parts and whole (see discussion in Craver & Bechtel, 2007). Bechtel and Richardson (2010: xxxiii) have taken “a step beyond decomposable and nearly decomposable modes of organization in characterizing *integrated* systems as those in which the operations of different component parts are interdependent; that is, they more or less continuously impact each other's operations”.⁶ Bechtel and Abrahamsen (2010) have called this new perspective *dynamic mechanistic explanation*: when the component parts of a mechanism are highly integrated, so that the behavior of a given part can be affected by the activity of many others, mechanistic explanations increasingly rely on mathematical modeling. Thus, in this perspective, the mechanistic models become more quantitative as will become conspicuous in Sect. 3.2, when I illustrate the MTE.

⁶The authors advocate that one of the simplest and more pervasive forms of interaction is *feedback*. For them, the important thing about mechanisms with feedback is the tension they place on the assumption of decomposability or near-decomposability. In their opinion, the more the various operations in the mechanism affect each other, the less successful is a sequential account of the mechanism in which each operation is treated as independent of the others. At an extreme there will be systems, or at least models of systems, in which the components are all uniform and the explanation of the resulting behavior appeals solely to the organization realized in the system. At such an extreme, the mechanistic heuristics of decomposition and localization may cease to be productive except insofar as their failure discloses functional integration. Bechtel and Richardson (2010) suggest that such behavior could be seen as “emergent” at least insofar as the organization of the system, rather than distinctive contributions of its constituent components, determines systemic function.

It is important to highlight the fact that Bechtel and Richardson (2010) reject a stricter and ruthless reductionism, in which lower levels are more important as a source of explanation.⁷ In this more nuanced position, the focus on the parts and operations must be relaxed and must shift toward the system through functional *recomposition*, in which one must show that the postulated component operations, with an appropriate organization, are sufficient to yield the systemic behavior. Therefore, Bechtel and Richardson (2010) assert the need to emphasize three points against ruthless reductionism within a mechanistic account: the fact that it is the whole organized mechanism that exhibits the phenomenon, not its component parts and operations; that mechanistic models are not typically governed by a single level;⁸ and that there are significant top-down constraints on the development of mechanistic models, in contrast with what is proposed by some mechanists that reject causal relations between different levels, as above referred. Thus, after localizing the parts and identifying the operations, a characterization of how the parts and operations are organized to make a functioning mechanism, and how the environment affects it, is needed. After decomposition and localization, the tasks of recomposing and situating the mechanism call for system thinking (Bechtel & Abrahamsen, 2005).

One of the major successes of the life sciences has been not just to identify numerous biological mechanisms but also to decompose them into their component parts and operations, with successful examples in different domains. Even though there are limitations in the basic account of mechanisms, I believe that one has to acknowledge that it describes the conceptual framework in which the vast majority of productive research in the life sciences has been conducted. However, and interestingly, philosophers of science have not exhaustively examined ecological

⁷According to Bechtel and Richardson (2010: xxxvii), implicit in their discussion of organization, complexity and emergence in mechanisms “is the basis for rejecting ruthless reductionism and for distinguishing mechanistic reduction from ruthless reduction or any other account that construes lower levels as the source of all explanation”. These authors also reject accounts that give a central role to laws in the explanation of complex systems, although acknowledging a limited role for laws in certain programs of mechanistic explanation. In their perspective, the parts and operations of a mechanism are organized, and the operations orchestrated in real time, such that the mechanism as a whole behaves in a particular way: it is not the operations alone, or even simply added together, that explain what the mechanism does; the specifics of their organization and their orchestrated execution must be addressed. Moreover, the parts and operations often are modified by other operations occurring elsewhere in the mechanism, and in some cases by activities external to the mechanism. That’s the reason why Bechtel and Richardson (2010) defend that the research must attend to the whole mechanism in its characteristic environment, and not just its lower-level parts.

⁸There is activity at each level, and neither level alone is sufficient to provide a complete account of the phenomenon in question. Appropriate tools are required for characterizing the ways in which the mechanism interacts with its environment as well as identifying its component parts and operations and their organization. It is also important to figure out how, as a result of the parts performing the operations they perform and these operations interacting in a well-orchestrated manner, the system exhibits the phenomenon when stimulated in particular ways by its environment. That is why, according to Bechtel and Richardson (2010), mechanistic explanations include, but go far beyond, identifying parts and operations at the lowest levels of organization.

mechanisms, even though ecologists describe mechanisms for purposes of explanation and prediction, whereby these latter might be different from the biological mechanisms that have so far received more philosophical scrutiny, such as the mechanisms uncovered in genetics, cell and molecular biology.

Ecologists have been intrigued since the beginning of the twentieth century by allometric and scaling relationships⁹ between organismal features and ecological phenomena (Brown et al., 2004; O'Connor et al., 2007; Raerinne, 2013, 2017). However, there was no widely accepted *mechanism* for the explanation of these patterns until recently. Very different explanations have been developed in the last two decades of the twentieth century, including those making reference to: structural and functional factors (e.g. surface area/volume effects on exchanges of heat and metabolites); biomechanical requirements for support and fracture resistance; natural selection on body size and life history (O'Connor et al., 2007). Thus, for some ecologists, the focus became on what different roles allometries and scaling relationships have in scientific mechanistic explanations and predictions in ecology, and how well allometries and scaling relationships are capable of functioning in these roles. Recently, several authors (West et al., 1997; Enquist et al., 2003; Brown et al., 2004; Allen & Gillooly, 2007; O'Connor et al., 2007) have developed a Metabolic Theory of Ecology (MTE) that, emphasizing allometric and scaling effects in ecology, has produced some interesting hypotheses.

After this brief introduction concerning the mechanistic account (Sect. 3.1), I shall therefore begin by characterizing MTE (Sect. 3.2). I then evaluate the virtues and limitations of this theory (Sect. 3.3). Following this evaluation, I explore and discuss the mechanistic nature of the theory (Sect. 3.4). Finally, I conclude by considering the general epistemological prospects of MTE for ecology (Sect. 3.5).

3.2 The Metabolic Theory of Ecology

MTE claims to provide a mechanistic explanation for known allometric relationships between biomasses and metabolic rates, postulating that these patterns of allometry and scaling relationships are driven by constraints of transport of energy and materials (West et al., 1997; Enquist et al., 2003; Brown et al., 2004; Allen & Gillooly, 2007; O'Connor et al., 2007).

This theory was advanced on the philosophical premises that allometric phenomena required a mechanistic explanation and that MTE could elucidate their mechanistic basis based on physical, chemical and physiological principles. More specifically, the formulation is based on the theoretical hypothesis that “the structure and dynamics of ecological communities are inextricably linked to individual metabolism” (Allen & Gillooly, 2007: 1073). This is a reductionist hypothesis in the

⁹In allometries and scaling relationships, body size (or mass) is used as an independent variable of different dependent variables representing anatomical, physiological, morphological, behavioral, social, ecological traits.

sense that interactions between individual organisms and their environment are constrained by metabolic rates, which in their turn depend on factors like body size, body temperature and resource availability; as a consequence, the interactions between individual organisms and their environment will explain the most significant characteristics of higher ecological levels.

Therefore, the reason for considering MTE a mechanistic approach is that metabolism is assumed to be mechanistically explained in terms of biochemical reactions. According to the basic mechanistic account that I have described in Sect. 3.1, decomposability is clearly possible in the case of MTE: the metabolism of a community (or ecosystem) is the behavior exhibited by the entire system (i.e., the whole mechanism), and the metabolisms of the populations that compose it are the components' activities of that entire system (i.e., the sub-mechanisms). In the same way, the metabolism of a population is the result of the metabolisms of the individual organisms, its components' activities. Thus, the levels of the mechanism are defined in terms of a relationship between the behavior of a system and the activities of component parts. That is, to say that a population is at a lower mechanistic level than the mechanism as a whole (ecosystem) is to say that it is a working part of the mechanism. The same is true for individual organisms being the working parts of a population.

Metabolism therefore provides a basis for using basic principles of physics, chemistry and biology to link the biology of individual organisms to the ecology of populations, communities and ecosystems. A calculation is said to be from basic principles if it starts directly at the level of established laws of physics. Thus, the ecology of populations, communities and ecosystems can be studied starting from the physical laws concerning the movement of gases, chemical elements, compounds and fluids, including the laws of diffusion and evaporation. Then, ecologists can use the laws involving the kinetics of chemical reactions. Finally, ecologists can analyze the variation in the rates and specificity of the biochemical pathways of metabolism among different kinds of organisms and environmental settings.

Metabolic rate i.e., the rate at which organisms take up, transform and expend energy and materials, is, according to this approach, the most fundamental biological rate. Advocates of this approach have developed a quantitative theory for how metabolic rate varies with body size and temperature (West et al., 1997; Enquist et al., 2003; Brown et al., 2004; Allen & Gillooly, 2007; O'Connor et al., 2007). This theory predicts how metabolic rate – by setting rates of resource uptake from the environment and resource allocation for survival, growth and reproduction – controls ecological processes at all levels of organization from individuals to ecosystems (Brown et al., 2004).

Within this perspective, the complex, spatially and temporally varying structures and dynamics of ecological systems are all consequences of individual metabolism because individual organisms transform energy to power their own activities, convert materials into uniquely organic forms, and thereby create a distinctive biological, chemical and physical environment. Metabolism might be generally characterized as the biological processing of energy and materials whereby organisms take up energetic and material resources from the environment, convert them

into other forms within their bodies, allocate them to the fitness-enhancing processes of survival, growth and reproduction, and excrete altered forms back to the environment. Metabolism therefore determines the demands that individual organisms place on their environment for all resources, and simultaneously sets powerful constraints on the allocation of resources to all components of fitness. In brief, the overall rate of these processes “sets the pace of life”, borrowing the expression of Brown et al. (2004: 1772). In particular, body size, temperature and chemical composition affect biological structure and function at various levels of organization (West et al., 1997; Enquist et al., 2003; Brown et al., 2004). This is because metabolism obeys physical and chemical basic principles that govern the transformation of energy and materials. Therefore, much of the variation among ecosystems – including their biological structures, chemical compositions, energy and material flows, population processes and species diversity – depends on the metabolic characteristics of the organisms living within them. The variation that is observed in individual organisms, including their life history, phenotypic features and ecological roles, is constrained by their own body sizes, operating temperatures and chemical composition (Brown et al., 2004). According to MTE advocates, these constraints of allometry, biochemical kinetics (i.e., rates of biochemical reactions) and chemical composition lead to “metabolic scaling relations that, on the one hand, can be explained in terms of well-established principles of biology, chemistry and physics and, on the other hand, can explain many emergent features of biological structure and dynamics at all levels of organization” (Brown et al., 2004: 1772).

MTE explicitly shows how many ecological structures and dynamics can be explained in terms of how body size, chemical kinetics and resource supply affect metabolism. MTE builds on them by providing a quantitative framework to better understand how these three variables combine to affect metabolic rate, and how metabolic rate, in turn, influences the ecology and evolution of populations, communities and ecosystems. The quantitative framework is illustrated by Brown et al. (2004: 1773–1786) as follows.

Allometries and scaling relationships can be represented as regression equations or power equations, in which one variable changes as a power of another.¹⁰ According to MTE, all characteristics of organisms vary predictably with body size, a correlation captured by the so-called allometric equations:

$$Y = aM^b \quad (3.1)$$

Y is the dependent variable, M the body mass, *a* is the normalization constant and *b* the allometric constant. In allometries and scaling relationships, body size or weight is treated as an independent variable of different anatomical, physiological, morphological, behavioral, social, ecological, and paleoecological dependent variables.

Depending on the value of their scaling exponents, allometries and scaling relationships are called either allometric ($b \neq 1$) or isometric ($b = 1$). Scaling exponents

¹⁰Regression analysis is a set of statistical techniques for estimating the relationships between a dependent variable and one or more independent variables.

can take both negative and positive values. In general, the larger the value of b , the faster Y increases (if b is positive in value) or decreases (if b is negative in value) with increasing M . If the scaling exponent, b , is less than unity, Y increases (or decreases if negative in value) more slowly than M does. On double log axes, the values of Y and M yield straight lines, and a gives the intercept or elevation of the regression line and b gives its slope. There is a plenitude of biological traits that correlate with body mass, M , and can be represented as dependent variables, Y . As Raerinne (2011: 194) illustrates, many scaling relations of putative ecological relevance can be derived on the basis of Eq. (3.1): fasting endurance scales as $aM^{0.44}$ for mammals and between $aM^{0.40}$ and $aM^{0.60}$ in birds; the size of the home range of birds and mammals varies positively with body size, aM^1 ; the inverse scaling rule: the maximum density, D , of herbivorous mammals declines as their body size increases, $D = aM^{-0.75}$; Kleiber's rule: basal metabolism, an estimate of the energy required by an organism for the basic processes of living, varies as $aM^{0.75}$; an individual's total energy consumption varies as $aM^{0.75}$; in most mammal groups gut volume is isometric to M , aM^1 ; heart rate varies as $aM^{-0.25}$.

Kleiber, in 1932, showed that the individual metabolic rate, I , scale as (Brown et al., 2004):

$$I = i_o M^{3/4} \quad (3.2)$$

The element i_o is the normalizing constant independent of body size.

It has been known that biochemical reactions rates, metabolic rates and nearly all other rates of biological activity increase exponentially with temperature. This kinetics is described by the Boltzmann factor or Vant' Hoff-Arrhenius relation (Brown et al., 2004):

$$e^{-E/kT} \quad (3.3)$$

E is the activation energy, k is the Boltzmann constant and T is the absolute temperature. This equation specifies how temperature affects the rate of reaction by changing the proportion of molecules with sufficient kinetic energy. This relationship holds over the temperature range of normal activity, which is 0–40 °C for most organisms.

Gillooly, in 2001, developed a model for the scaling of metabolic rate that combines the effects of body size and temperature, preliminarily conceptualized in Eqs. (3.1), (3.2) and (3.3), which leads to a single equation for individual metabolic rate, I (Brown et al., 2004; Allen & Gillooly, 2007):

$$I = i_o M^{3/4} e^{-E/kT} \quad (3.4)$$

This simple analytical expression yields quantitative predictions on metabolic rate that are supported by empirical data for a broad assortment of taxonomical groups. As a result, its explanatory power has been claimed to be substantial (Brown et al., 2004; Allen & Gillooly, 2007).

Characteristics of organisms vary with their body size, temperature and chemical composition or *stoichiometry*. In ecology, stoichiometry refers to the quantities, proportions or ratios of chemical elements in different entities, such as organisms or their environments. All organisms have internal chemical compositions that differ from those in their environments; therefore, they must expend energy to maintain concentration gradients across their boundaries, to acquire necessary elements and to excrete waste products. Fundamental stoichiometric relations dictate the quantities of elements that are transformed in the reactions of metabolism. Biochemistry and physiology specify the quantitative relation between the metabolic rate and the flows of elements through an organism. The metabolic rate dictates the rates at which material resources are taken up from the environment, are used to enact biological structure and function, and are finally excreted as waste back to the environment. The chemical equations of metabolism specify not only the molecular ratios of elements, but also the energy yield or the demand of each reaction. Ecological stoichiometry is concerned with the causes and consequences of variation in elements' composition among organisms and between the organisms and their environment; indeed, there is great variation within and among organisms, and especially between different taxonomic or functional groups. The concentrations of elements in ecosystems are therefore directly linked to the flows and turnover rates of elements in the constituent organisms (West et al., 1997; Enquist et al., 2003; Brown et al., 2004; Allen & Gillooly, 2007). On the one hand, environmental concentrations can limit metabolic rates, and thereby growth rates, reproductive rates and quantities of organisms; on the other hand, the size of stock of elements and rates of turnover in organisms can regulate environmental concentrations of elements and compounds. The Eq. (3.4) gives, as referred above, the combined effect of body size and temperature on individual metabolic rate, I . Because the mass specific rate of metabolism, B , is simply I/M , it follows that B scales as:

$$B \propto M^{-1/4} e^{-E/kT} \quad (3.5)$$

The advantage of this concise mathematical expression is that it combines the effect of body size and temperature in a single quantitative expression. This allows making precise comparisons between organisms and different functional and taxonomical groups differing substantially in these variables (Brown et al., 2004). When such comparisons are made, the commonalities of life and their ecological manifestations are revealed. Brown et al. (2004: 1775) present empirical data plotting rates of T-corrected individual production against body mass for a wide variety of organisms, showing that MTE predicts that Eq. (3.5) should account for much of the variation in some characteristics of individual performance and life history, such as individual biomass production, ontogenetic growth, survival and mortality, as well as stoichiometry.

MTE extends this framework to population and community levels of ecological organization, claiming that many features of population dynamics and community organization are due to effects of body size, temperature and stoichiometry on the performance of individual organisms. The maximal rate of exponential increase in

a population, r_{\max} , is predicted to scale according to Eq. (3.5) (Brown et al., 2004); this inference follows from the fact that reproduction is fueled by metabolism, and that mass-specific production rates and mortality rates follow the same equation. Advocates of this metabolic ecological approach contend that it is possible to explain the equilibrium number of individuals or carrying capacity, K , predicted to vary as:

$$K \propto [R] M^{-3/4} e^{-E/kT} \quad (3.6)$$

Therefore, K varies linearly with the supply rate or concentration of the limiting resource $[R]$, as a power function of body mass and exponentially with temperature. Thus, if $[R]$ increases, there will be more organisms of decreased size. However, if temperature increases, the carrying capacity is reduced because the same supply of energy supports a smaller number of organisms, each fluxing energy and materials at a higher rate. Notably, Brown et al. (2004: 1775) present empirical evidence for an inverse Boltzmann relationship between equilibrium abundance and environmental temperature.

The mathematical structure of the theory, according to its authors, thus provides two concise mathematical expressions of particular interest, derived from Eq. (3.4): Eqs. (3.5) and (3.6). The first correlates the variation of B , the mass specific rate of metabolism,¹¹ with the combined effect of size and temperature in a single quantitative expression. What is observed is that B is negatively associated with body size and positively (exponentially) with temperature. The second correlates the variation of K , the carrying capacity of individual organisms,¹² with the combined effect of size, temperature and the available quantity of a limiting resource. What is observed is that the carrying capacity is negatively associated with body size, and positively with the resource concentration (linearly) and temperature (exponentially).

Ecologists have tried to understand how pairs of competing species or of predators and preys stably coexist in the same environment. Empirical evidence suggests that a number of interaction rates and times, including rates of parasitism and predator attacks, are inversely related with temperature. It has been argued that MTE predicts the pace of these interspecific interactions, because the rates of consumption and population growth are determined by the rates of individual metabolism and have the same body size and temperature dependence (Brown et al., 2004; Allen & Gillooly, 2007). Moreover, the scaling of rates of ecological interactions has important implications for coexistence and species diversity, because the qualitative empirical patterns of biodiversity would suggest that the processes that generate and

¹¹Mass specific rate of metabolism is the rate at which organisms consume energy per unit of body weight.

¹²The carrying capacity of an environment or given place is the maximum population size of a biological species that can be sustained by that specific environment, given the food, habitat, water, and other resources available. The carrying capacity is defined as the environment's maximal load, which in population ecology corresponds to the population equilibrium, when the number of deaths in a population equals the number of births.

maintain species richness scale similarly to other biological rates, as illustrated by Eq. (3.5). Other things being equal, there are more species of small organisms than larger ones as well as more species in warm environments than colder ones.

The corroborated hypothesis that species diversity varies inversely with body size suggests, according to MTE, that metabolism plays a central causal role in determining ecosystems' species composition. It has long been known that the diversity of most taxonomic and functional groups is highest in the tropics, but this has usually been attributed to higher productivity or reduced seasonality, rather than to the kinetic effect of higher temperature (Pianka, 1966).¹³ However, empirical evidence suggests that species richness in many groups of animals and plants has the same relationship to environmental temperature that metabolic rate has (Brown et al., 2004). This result holds true not only along latitudinal gradients, but also along elevation gradients, where variables such as light intensity, seasonal changes in day length and biogeographic history are held relatively constant. The implication is that much of the variation in species diversity is directly attributable to the kinetics of biochemical reactions and ecological interactions (Brown et al., 2004; Allen & Gillooly, 2007). Clearly, much additional work on the relationship between metabolism and biodiversity is needed, but a metabolic perspective, as proposed by MTE, demonstrates the centrality of many of these questions and suggests ways to look for in pursuit of appropriate answers.

Some of these questions can be addressed by assessing the effects of biological metabolism on the paths of energy and materials in ecosystems (Brown et al., 2004: 1782). It may be suggested that the biologically regulated whole ecosystems' stores and fluxes of elements and compounds are simply the *sums* of the stores and fluxes of the constituent organisms. Thus, MTE is putatively able to predict the contribution of the biota to biogeochemical cycles. Specifically, in this framework, Eq. (3.6) provides the ground for predicting how body size, temperature and stoichiometry determine specific magnitudes of stores and rates of flux within and between "trophic compartments" such as primary producers, herbivores, predators and detritivores.¹⁴ It is possible to derive from Eq. (3.6) expressions for the stored biomass, the energy flux and biomass production, the biomass turnover, and for trophic dynamics (Brown et al., 2004: 1784). MTE also provides a framework for more explicitly incorporating stoichiometry and understanding the effects of limited water and nutrients supply on variation in productivity and other processes across biomes (i.e., collections of organisms that have common characteristics with respect to the

¹³Pianka's (1966) paper is, according to most authors, the first attempt to explain mechanistically some ecological phenomena.

¹⁴Primary producers, otherwise known as autotrophs, are organisms that convert energy (through the process of photosynthesis) into food. Herbivores are organisms that principally eat autotrophs such as plants, algae and photosynthesizing bacteria – more generally, organisms that feed on autotrophs in general are known as primary consumers. Predators are organisms that use predation for feeding (normally on herbivores). Detritivores, otherwise known as decomposers, are organisms that break down chemical compounds from dead bodies of producers and consumers into simpler forms that can be reused.

environment they inhabit) and physical gradients. According to the MTE advocates, regressions incorporating these variables are able to account for much of the observed variation.¹⁵

In summary, MTE:

1. conjectures that a complex structure of distributional networks of essential nutrients and chemical elements, such as the circulatory systems in metazoans (i.e. multicellular animals), requires an allometry in order to explain the need to minimize transport costs of energy and materials as body size increases;
2. hypothesizes that minimizing these transport costs requires a scaling exponent of $-3/4$, as defined in Eq. (3.5);
3. links metabolism and temperature, via the Boltzmann factor, used to predict the rate of simple biochemical reactions, essential to living processes, as defined in Eqs. (3.5) and (3.6).

The theory was advanced on two philosophical premises (O'Connor et al., 2007: 1059). First, that allometric phenomena *require* a mechanistic explanation and that MTE is able to *identify* that mechanism, based on physico-chemical, biochemical and physiological basic principles. That is, MTE characterizes a mechanism, according to the account described in Sect. 3.1: the metabolism of an ecosystem (or community) is the behavior of the mechanism as a whole, the *phenomenon* to be explained, the relevant component *parts* are the individual organisms composing the populations, which in their turn compose communities, which in their turn compose ecosystems; the metabolisms of populations (and the metabolisms of individual organisms) are the relevant *operations* of the component *parts*, *organized* in accordance to the levels of ecological organization and *causally* related to produce the phenomenon. The second premise is that the postulated mechanism motivates and justifies renewed and expanded work on the ecological implications of allometry. I think these premises are valid, with the caveat that allometries and scaling relations are not universal or exceptionless laws; on the contrary, they capture observable tendencies, generalizations underpinning ecological systems (Lawton, 1996, 1999; Colyvan, 2003). If it is true that allometries and scaling relationships do not represent biological laws, the covering-law account (Hempel, 1965) cannot be used to explicate how and under what conditions they function in articulating explanations and inferring predictions. Thus, in order to salvage the putative explanatory and predictive roles of allometries and scaling relationships, one feasible alternative

¹⁵In statistics, the coefficient of determination, denoted as r^2 , is the proportion of the variance in the dependent variable that is predictable from the independent variable(s). It is a statistical measure used in the context of statistical models whose main purpose is either the prediction of future outcomes or the testing of hypotheses, on the basis of other related information. It provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. Thus, in the literature that explores this kind of correlation between body mass as an independent variable and other traits as dependent variables, it captures how the correlation is dependent on the indices of fit, usually the value of r^2 . The advocates of MTE argue that the coefficient of determination presents good values in most of the empirical examples.

is that provided by a mechanistic approach because, as I have related in Sect. 3.1, in such account mechanisms are contrasted explicitly with laws of nature (Machamer et al., 2000; Craver & Tabery, 2019).

The putatively mechanistic explanation articulated by MTE has energized the consideration of allometric effects in ecology, producing a number of substantial hypotheses. Consequently, the expectation was that MTE would be able to provide a quantitative framework to better understand how the variables of resource availability, body size and temperature combine to affect metabolic rate and, in turn, how metabolic rate influences the ecology of populations, communities and ecosystems. However, there is a considerable debate regarding the support for the predictions articulated by MTE as well as the validity of the theory's underlying assumptions. This is the issue that I shall explore in the next section.

3.3 Virtues and Limitations of MTE

O'Connor et al. (2007) recognized two putative advantages of MTE: the first is that it is based on basic principles of physics, chemistry and biology; the second is that it depends on fewer assumptions and parameters than other explanatory frameworks. However, these authors are also of the opinion that the notion of first principles should be defined in a better way. Additionally, they argue that relative freedom from assumptions is hardly an advantage in itself. In fact, while well-elaborated ecological models based on physical laws and relationships have the virtue of, first, depending on models of processes whose dominant dynamics are putatively better understood and, additionally, of clearly outlining the models' underlying assumptions, they are not free of such assumptions. In fact, the application of simple physical principles to ecological systems requires many assumptions because of the complex series of organizational levels, ranging from macromolecules to ecosystems, through which physical effects are filtered to produce ecological dynamics. O'Connor et al. (2007: 1060) contend that the potentially limiting assumptions of MTE regarding metabolic allometry include: that metabolic rate is primarily limited by distributional networks of nutrients (such as circulatory systems); that circulatory transport costs will indeed be minimized; that the capillary diameters of such networks are size invariant; that the simplified description of the circulatory system is inadequate, even for organisms with open circulatory systems or no cardiovascular systems; that branching in real bodies sufficiently approximates the simplifying assumptions to justify MTE arguments; and that the normalization constants in the allometric power equations are unimportant in comparing scaling exponents.

Thus, following O'Connor et al.'s (2007) argument, MTE shall be hardly free from or independent on fewer assumptions. I shall now argue that some of these are hardly tenable. Characterizing MTE as based on first principles does not mean that

the assumptions are easily identifiable and the consequences of violating them are easily understood. Moreover, all this should not confer special status on MTE as an explanatory mechanism.

In spite of these possible limitations, some authors disagree with O'Connor et al. (2007), arguing that the proposed mechanisms underlying the body size and temperature dependencies of individual metabolic rate represented in Eq. (3.5) can be accurately described (Allen & Gillooly, 2007: 1074). O'Connor et al. take issue with the Brown et al.'s network model as a mechanistic explanation for the $3/4$ power scaling of metabolic rate (2004); however, Allen and Gillooly (2007: 1075) argue that this model is indeed mechanistic because: it invokes a few simplifying assumptions that allow causes to be postulated; it yields quantitative predictions by explicitly linking organism structure to function based on these assumptions; it can be extended to predict how deviations from assumptions affect model predictions. This network model represents a manifestation of general principles that entail the maximization of the number of metabolic units where metabolism occurs – as respiratory complexes, for instance – and, at the same time, minimizing the transport distances to those metabolic units (Allen & Gillooly, 2007: 1074). These general principles are geometric. They are applied by assuming that natural selection will lead to evolutionary optimization of network geometry, subject to physical and physiological constraints. Given this evolutionary optimization assumption, quarter-power scaling of metabolic rate is predicted to apply at multiple levels of biological organization, in agreement with what is showed by the empirical data (Allen & Gillooly, 2007). O'Connor et al. (2007: 1061–1062) criticize the general hypothesis that natural selection results in the network optimization in organisms. According to these authors, the minimization of transport costs is a tenable criterion for evolutionary optimization, although it is clearly not the only factor upon which selection operates; consequently, this simple optimization criterion is necessary but not sufficient to describe the variety of selective forces that likely operate in determining metabolic rates. Thus, for O'Connor et al. (2007: 1062), the degree to which minimizing a subset of costs of metabolite transport determines metabolic rates needs to be substantiated by further research. In the perspective of these authors, to whatever extent selection might optimize bulk transport via distributional networks, the specific mechanism proposed by the MTE (i.e., isolated minimization of fluid transport costs, without taking into account other criteria) is an insufficient explanation for minimization because it is unlikely that transport costs map in a uniform and simple manner onto fitness. In my view, O'Connor et al. (2007)'s stance is correct, because natural selection optimizes organismal fitness (and not the fitness of isolated organismal parts), implying that the optimal design hypotheses must consider costs and benefits of parts optimization (O'Connor et al., 2007). As these authors say, "what is actually required in order to maximize fitness is that simultaneous optimization, of both the costs and benefits, is expressed in a common currency, organismal fitness being the logical candidate" (O'Connor et al., 2007: 1062).

In my opinion, this is the correct perspective: true optimization of isolated physiological systems (like metabolism) is difficult, if not impossible, to achieve in nature. Pleiotropy,¹⁶ multiple use for structures (in the sense of a metabolic component part being causally involved in non-metabolic processes), and variable selective environments all constrain optimization of physiological systems. Selection more likely might optimize the reproductive success of entire organisms rather than the efficiency of a particular component of metabolism, although the two may well be correlated. Many aspects of metabolism, active and resting, could conceivably heavily affect the fitness of an organism. Therefore, the idea that natural selection would optimize circulation transport costs (in isolation from other systems), which would then come to dominate metabolic allometry, seems unlikely. When conflicting demands are placed on a system, biologically optimal solutions are context dependent and the results of optimality analysis depend on the optimization criteria. Rarely can any of a set of criteria be confidently identified with the fitness of an organism. This is the issue raised, as I have anticipated in the introduction, by the supposed near-decomposability of structures into sub-structures, piece by piece. O'Connor et al. (2007: 1063) present evidence that shows that predictions made by MTE regarding single cells are poorly supported by data and that empirical evidence disputing the main predictions of MTE with regards to the mechanism of metabolic standing and transport systems in vascular plants is growing. According to these authors, it is difficult to understand how a cell of a multicellular organism, being part of the whole organism, can be optimized in an isolated way, i.e., expecting that component parts' optimization ensues and furthermore causes organism optimization in terms of transport distances. However, Allen and Gillooly (2007: 1074), opposing this criticism, argue that there is considerable empirical evidence to support the network model assumptions and predictions in unicellular organisms, plants and a variety of animal taxa.

Nevertheless, even if I concur with O'Connor et al.'s (2007) criticisms based, first, on the difficulties related to decomposability and, secondly, on picking out the relevant structural/functional units of the mechanism underlying MTE concerning the optimization of nutrient circulation, I suggest that Bechtel's (2015) approach – in which living systems are considered to exhibit the properties of *scale-free small-world networks* can be valuable on this issue. The idea is to draw boundaries for explanatory purposes at particular locations. In such networks, some nodes can have a number of connections much larger than others, in which case there is no scale for characterizing the distribution of number of connections. This is why these networks are termed scale-free (Bechtel, 2015: 89). The issue here is that there is no point on the scale at which one can capture all the relevant connectivity. This can be problematic since the nodes with more connections are typically very important to

¹⁶Pleiotropy occurs when one gene influences two or more seemingly unrelated phenotypic traits. A gene exhibiting multiple phenotypic expression is called a pleiotropic gene. Mutation in a pleiotropic gene may have an effect on several traits simultaneously, due to the gene coding for a product used differently by a population of cells or having different functions in different tissues.

the functioning of the network, due to the fact they have more widespread effects, and for this reason O'Connor et al. (2007) criticize predictions for transport systems based on single cells. Maybe the modularity that I have referred to in Sect. 3.1 can be helpful, because in Bechtel's (2015) perspective a module in a scale-free small-world network can be considered as a mechanism if its nodes collaborate in the production of a phenomenon. Considering the case of a circulatory system, to take fully into account how the module behaves, a mechanist would need to take into account both the connections between the nodes in the module and the connections between the same nodes and the nodes elsewhere in the larger network (e.g., respiratory system, digestive system). Therefore, the mechanistic account would only focus on the connections within the module, except for connections that are treated as providing inputs and outputs (Bechtel, 2015). This would be a misrepresentation of how the module behaves, since it does not consider all the interactions with nodes outside the module, a critique in line with O'Connor et al.'s (2007). However, I believe that Allen and Gillooly (2007) think that the proposed mechanism of MTE provides an account that approximates the behavior of the mechanism, and that can be "accurate to a first approximation" (Bechtel, 2015: 89). Despite these promising developments within New Mechanistic Philosophy, it seems to me that, even recognizing that the alternative would be a holism making it impossible to identify the distinctive contribution of the parts of the system, mechanist researchers will often discover that the mechanisms they investigate are much more integrated than they initially had assumed.

Another criticism advanced against the MTE by O'Connor et al. (2007: 1063), and with which I partly concur, is that, even accepting that most components of metabolism are affected by temperature, proposing Boltzmann relationships as the only explanatory mechanism for the temperature dependence of metabolism is untenable because no macroscopic equivalent of activation energy exists.¹⁷ At the level of the cell, with its complex, feedback controlled network of reactions with numerous metabolic checkpoints, each of which responds to several controllers and external conditions, no equivalent of activation shall exist. Furthermore, at this level the simplicity of control implied by the Boltzmann relationship does not exist (O'Connor et al., 2007). Accordingly, as one moves to consider organ systems, entire individual organisms, populations and ecosystem levels of organization, the complexity and diversity of the organization networks and their responses to temperature all increase progressively. Any putative relationship based on activation

¹⁷In chemistry and physics, activation energy is the minimum amount of energy that must be provided for compounds to result in a chemical reaction. Activation energy can be thought of as the magnitude of the potential barrier (sometimes called the energy barrier) separating minima of the potential energy surface pertaining to the initial and final thermodynamic state. For a chemical reaction to proceed at a reasonable rate, the temperature of the system should be high enough such that there is an appreciable number of molecules with translational energy equal to or greater than the activation energy. Therefore, and following the argument, the Boltzmann relationship can only offer at a micro-level, and very locally, an explanation for the temperature dependence of metabolism, and not at a macro-level.

energies for the processes of the different levels of organization can be, at best, an observed correlation with temperature, but not necessarily a mechanism. Mechanisms demand productive causation adequate to describe phenomena, both qualitatively and quantitatively, that is, a connection between the dependent effect and the causative process. In this respect, O'Connor et al. (2007) have a substantive criticism: it would be very difficult to identify a causative process that could explain the effect of temperature moving across all the levels of ecological organization. Moreover, as I have related in the introduction, the near decomposability of mechanisms is directly related to the idea that mechanisms span multiple levels of organization. However, this view, according to some mechanists, has the implication that there can be no causal relationships between items at different levels of mechanisms (Craver & Bechtel, 2007; Craver & Tabery, 2019), making it difficult to postulate the role of a causative process moving across all the ecological levels. However, if we adopt a less ruthless mechanistic account, such as that advocated by Bechtel and Richardson (2010), causal relations between different levels can be recognized, and the effect of temperature moving across all the levels of ecological organization could be possibly explained. In this sense, I would argue that Bechtel and Richardson's stance needs further attention, and maybe MTE can elaborate on this issue along similar lines, coming up with a suitable answer. Probably most importantly, according to O'Connor et al. (2007: 1063) correlative relationships like those proposed by MTE cannot be regarded as mechanistic because they are merely couched in a mathematical form, lacking the reference to the causative linkage between the postulated mechanism and the temperature dependence of metabolism. O'Connor et al. (2007: 1063) argue that the reservation about Boltzmann's relation arises from "a poor fit to thermal physiology, particularly the short- and long-term variation of metabolism in response to changes in body temperature". O'Connor et al. (2007: 1064) present data from the literature on acclimation,¹⁸ particularly acclimation of metabolic rates to varying temperatures in ectotherms,¹⁹ showing that the dependence of metabolism on temperature is both subject to selection among animal taxa and physiologically adjustable within a single organism. Thus, this range of variation is inconsistent with Boltzmann's relation. O'Connor et al. (2007) do not argue that temperature does not affect metabolism, but that the proposed Boltzmann relationship cannot wholly encompass the patterns of variation commonly seen in nature. On the contrary, Allen and Gillooly (2007: 1075) argue that the Boltzmann relationship can capture the complexities of metabolism, claiming that a large and growing body of empirical evidence supports the commonality of temperature responses rather than the ability of individual species to overcome

¹⁸Acclimation is the process in which an individual organism adjusts to a change in its environment (such as a change in altitude, temperature, humidity or pH), allowing it to maintain performance across a range of environmental conditions.

¹⁹An ectotherm is a type of organism in which internal physiological sources of heat are of relatively small or of quite negligible importance in controlling body temperature. Such organisms (for example frogs) rely on environmental heat sources, which permit them to operate at very economical metabolic rates.

the physical constraints of temperature. These authors argue that the Boltzmann's relation is firmly based in statistical thermodynamics and present empirical data relative to heterotrophic organisms (diverse taxa of insects and marine larvae), showing that the temperature dependence of metabolic rate reflects the temperature dependence of respiration on individual mitochondria, according to Eq. (3.5). They also show that, for plants, this same temperature dependence is expected to hold over the short term.

Bechtel and Bollhagen (2021) try another approach to activities in biological mechanisms. These authors start from the acknowledgement that, on the standard account of mechanistic explanation (outlined in Sect. 3.1), a phenomenon is explained by appealing to the entities and activities constituting it, even though "the active nature of activities remains unexplained" (Bechtel & Bollhagen, 2021: 12721). Following the same kind of reasoning, the standard account is not sufficient to understand how mechanisms are active. In the MTE case, we should be able to explain why there is a continuous metabolism in individuals, populations, communities and ecosystems, being the source of activity the metabolism of individual organisms. Accordingly, one should recognize that mechanisms are only active when free energy is employed by them,²⁰ something that has not been emphasized in the accounts of New Mechanists, but that is fundamental because a mechanism without free energy will not perform work. What happens to the free energy depends on how it is *constrained* because, if it is not constrained, it simply dissipates, with an increase in entropy and without the possibility of generating work; in contrast, if it is constrained, work can be performed, with the nature of the work depending on the constraints imposed (Bechtel & Bollhagen, 2021). What is then assumed is that there are entities engaging in activities, whose basic explanatory principles are *energetics* and *constraints*, showing that there is a need to clarify what the activities are in order to explain the mechanism, subjecting them to further analysis (Bechtel & Bollhagen, 2021: 12723). When Allen and Gillooly (2007) argue that the Boltzmann relationship can capture the complexities of metabolism in MTE, and that there is a common response to temperature, I would argue that these authors, for the purpose of explaining the activities of the entities involved (individual organisms, populations, communities and ecosystems), are construing them based on the activity underlying organismal metabolism being the reference, which is constrained in different manners (because different species have different metabolisms). That is the reason why the activities of different entities respond in distinct characteristic ways as free energy flows through them, and the Boltzmann relationship is able to generalize over this variation.²¹ Bechtel and Bollhagen (2021) also argue that energetics

²⁰In thermodynamics, the Gibbs free energy (or Gibbs energy) is a thermodynamic potential that can be used to calculate the maximum reversible work that may be performed by a thermodynamic system at a constant temperature and pressure.

²¹The Gibbs free energy, ΔG , ($\Delta G = \Delta H - T\Delta S$, where H is enthalpy, S is entropy, and T is temperature, measured in joules in SI) depends on temperature, such as metabolism depends, following the capture of the Boltzmann's relation in the allometric equations proposed by MTE, (see Sect. 3.2).

and constraints are also relevant to understanding mechanisms at higher levels of organization, and across all levels in the mechanistic hierarchy. Just as the bottom-level, higher-level activities depend upon the release of energy and, importantly, higher-level entities also constrain those at the bottom level, determining, for instance, how energy used and released by individual metabolism results in the activities of populations, or even communities and ecosystems. This fits well with Bechtel and Abrahamsen (2010) perspective of a *dynamic mechanistic explanation*, when the parts of a mechanism are highly integrated, as I have described in Sect. 3.1. These mechanistic developments can benefit MTE, in the sense that they can help to localize where the work is done, ultimately explaining the source of the activity of the relevant component parts of the mechanism, providing a reference point for understanding the operation of the whole mechanism, that, in the case of MTE, would be the ecosystem.

Another point of contention is that, according to O'Connor et al. (2007: 1061), these attempts to link individual metabolic rate to the structure and function of higher levels of biological organization are neither useful nor valid: MTE only describes “pre-existing” patterns, and these patterns are dissociated from underlying mechanisms. Allen and Gillooly (2007: 1075) argue that these claims are false and reflect a poor understanding of MTE; in fact, these authors argue, “the vast majority of MTE studies have been motivated by new questions that have resulted in the generation of new hypotheses, models, and empirical relationships”. In each of these studies, new patterns have been described and directly linked to individual metabolic rate.

In this section, I have showed that a significant debate about the capacity of MTE to explain and predict ecological phenomena has taken place since its inception. I shall now discuss the issue of the mechanistic nature of MTE because a central question needing clarification is whether the explanation in terms of allometric relationships provided by MTE is actually mechanistic.

3.4 The Mechanistic Nature of MTE

In the discussion concerning the mechanistic nature of MTE, it is crucial to distinguish between different types of explanatory models (O'Connor et al., 2007; Pâslaru, 2009; Raerinne, 2011; Raerinne, 2013): a *phenomenological* type, which requires an empirical, but not a mechanistic, relation between variables – e.g., regressions of metabolic rate on mass fall in this category –, where this implies that perhaps there are hidden variables creating the pattern, thus engendering a problem of extrapolation and prediction; a *mechanistic* type, in which the input variables are indeed linked to output variables by a series of causal relationships, whereby these variables can be considered mechanistic. The models proposed by MTE are, I argue,

a compromise between these two types of models; in fact, proxy variables²² for factors of likely mechanistic importance are embedded within MTE's statistical framework.

In spite of the long and useful history of the program of physiological explanation of ecological processes, the extent to which regressions provide broadly generalizing mechanisms for variable physiologies, distributional systems of chemical elements and their constraints, and selection forces, remains an open question. Likewise, in my opinion, the extent to which the linear equations,²³ fitted by regression, constrain the mathematical form of a generalized mechanism is unclear, particularly in systems whose dynamics result from multiple, interacting, nonlinear subsystems such as ecosystems.²⁴ Therefore, some critics of MTE argue that most putative MTE mechanisms must be regarded as mere statistical models based on plausibly mechanistic variables. According to O'Connor et al. (2007: 1060), "arbitrarily assigning some components of interspecific allometric variation to the scaling, while assigning others to the normalization constant, is a reification of

²²In statistics, a proxy is a variable that is not in itself directly causally relevant, but that serves an epistemic function in place of an unobservable or immeasurable variable. In order for a variable to be a good proxy, it must have a close correlation, not necessarily linear, with the variable of interest. This correlation might be either positive or negative. Raerinne (2013: 195) exemplifies this with the following scenario. In antelope species, body size is positively correlated with group size, and the two variables covary from small, almost solitary antelope species, such as the species in the genus *Cephalophus*, to large herding species, such as *Oryx beisa*. A qualitatively similar correlation between body size and group size has been found in primates too. In the regression for antelope species, however, body size is a proxy for or a correlate of other variables, such as the available food supply in the home range of these species or the level of predation threat. The latter two, rather than body size, seem to represent the true causes or explanatory factors for differences in group sizes among antelope species. Body size is therefore used as an independent variable for *convenience*: it is more easily quantifiable and measurable than the putative true causes of the phenomenon. For instance, species' home ranges can be notoriously difficult to estimate, let alone to measure. Presenting a proxy as a cause is not only inaccurate but also misleading insofar as we are searching for ways to control and understand nature. Of course, the use of proxies is unobjectionable if the authors acknowledge this limitation and do not present proxies as causal factors. However, proxies are sometimes interpreted causally because the true causes or effects are not easily definable as variables, in which case, proxies can be used to hide the fact that the allometries and scaling relationships in question are non-change-relating as generalizations (see below).

²³In **mathematics**, a linear equation is an **equation** that may be put in the form: $a_1x_1 + \dots + a_nx_n + b = 0$, where x_1, \dots, x_n are the variables, and b, a_1, \dots, a_n are the coefficients, which are often **real numbers**. The coefficients may be considered as **parameters** of the equation, and may be arbitrary **expressions**, provided they do not contain any of the variables. To yield a meaningful equation, the coefficients a_1, \dots, a_n are required not to be all zero.

²⁴In **mathematics**, a nonlinear system is a **system** in which the change of the output is not **proportional** to the change of the input. Nonlinear **dynamical systems**, describing changes in variables over time, may appear chaotic, unpredictable, or counterintuitive, contrasting with much simpler **linear systems**. Typically, the behavior of a nonlinear system is described in mathematics by a nonlinear system of equations. The point of interest, in my opinion, is that higher ecological levels, such as communities and ecosystems, behave as nonlinear systems – thus, one cannot fit complex ecological data in a linear regression (which uses linear equations), even using plausible mechanistic variables such as temperature, for example.

assumptions of statistical fitting and not an expression of mechanistic understanding”. Such speculations may ultimately be justified by the discovery of genuine mechanisms, but so far they are merely statistical and, as such, they do not possess the robustness of mechanisms and cannot be justified by the very patterns to which the statistical models fit.

In this sense, I would argue that it is difficult to understand how allometries and scaling laws, such as those exemplified by West et al. (1997), Enquist et al. (2003), and Brown et al. (2004), represent generalizations with causal or explanatory relevance for ecology, in the sense of representing a causal or explanatory relation between variables, dependent and independent, with well-defined values. Even though it is well known that correlation does not necessarily amount to causation or provide causal explanation, in practice this point is often forgotten in the literature on allometries and scaling relationships, whereby body size or mass as an independent variable is claimed to explain a major part of the variation in the dependent variable. In this respect, it is instrumental to consider the analysis of causation suggested by Raerinne (2011), who defends an interventionist account of causal explanation to which I have already referred to in the introduction (see footnote 4). In such an account, *invariance* should be the correct relation of explanatory relevance in the case of causal explanations, which should be conceptualized as descriptions of objective dependency relations between entities or variables. Raerinne (2011: 253) claims that an invariant generalization “is one that continues to hold under a special change – *an intervention* – that alters the value of its variables”. According to this interventionist account of explanation, regressions or correlations,²⁵ by themselves, are not explanatory, regardless of how strong the correlation between the two factors is. For a correlation to count as genuinely explanatory, an intervention must be performed, whereby the relationship between factors actually remains invariant. In order for there to be an intervention and a possibility of manipulation, at least some of the predicate terms of a generalization are required to be representable as variables. If a generalization cannot be tested for how it might behave under interventions in or manipulations of its variables, the claims made about its explanatory status should be parsimonious.

Accordingly, many large-scale ecological generalizations are not explanatory for the reason that they do not describe invariant relations. Even though some of these relationships represent change-relating generalizations,²⁶ it is quite possible that they might be joint correlation effects, because of their common causes or the causal influence of certain background conditions. That is, they would amount to cases of “spurious causation”, in the words of Raerinne (2013: 195). Moreover, even if one

²⁵In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense correlation is any statistical association, though it commonly refers to the degree to which a pair of variables is linearly related.

²⁶A change-relating generalization describes how changes in the value of its variable or variables are related to the changes in the value of other of its variables. Change-relating generalizations typically describe dynamic or active relationships between variables.

finds that some of these generalizations are change-relating and invariant, allometries and scaling relationships seem to rather offer phenomenological explanations that require, in order to be genuinely explanatory, to be supplemented with information about the *mechanisms* underlying them. Allometries and scaling relationships with ill-defined variables are thus non-explanatory as generalizations. Indeed, in ecology, phenomenological explanations – in which one has an invariant relation between variables but no account as to why or how the relation holds between variables – are abundant (Raerinne, 2011).

In the interventionist account of explanation defended by Raerinne (2011), causes are *difference-makers*. Causes and effects should be understood as representable variables. Causes are difference-makers in the sense that they can be intervened upon to control or manipulate their effects. A change in the value of a cause makes a difference in the value of its effect. It is useful to distinguish between two kinds of causal explanation in the philosophical literature: *simple causal claims* and *mechanistic explanations* (Raerinne, 2011, 2013). A simple causal claim describes the causal connection between the phenomenon to be explained and the thing that does the explaining. It refers to a *phenomenological* or *superficial* causal explanation in which one has an invariant relation between variables, but no account – or mechanistic explanation – as to why or how the relation holds between variables. This account of explanation describes how simple causal claims function by identifying what is required of a causal dependency relation in order to be considered explanatory. That is, simple causal claims need to be invariant during interventions. Describing a mechanism of a phenomenon is not something that is contrary to the spirit of describing what the causal dependency relation of a simple causal claim actually is. Instead, a mechanistic explanation is a complement to a simple causal claim, since it describes *how* the dependency relations produce the phenomenon to be explained. In particular, a mechanistic explanation describes the internal causal structure of the phenomenon to be explained, as I have related in Sect. 3.1. It describes the underlying mechanism *within* the system by showing how the system is constituted and how the mechanism produces the phenomenon to be explained. According to Raerinne (2011), mechanistic explanations are causal and bottom-up reductionist explanations; they are causal explanations as a result of their invariance; they are true if they correctly describe the mechanisms in nature.

Describing an “underlying mechanism” becomes, henceforth, a paramount complement to an invariant causal relationship because it shows *how* the relationship produces the phenomenon (Raerinne, 2013). A causal explanation that is complemented with mechanistic details provides us with possibilities of more precise interventions and information about the extrapolability of a causal relationship, because with mechanistic details we obtain information about how the parts of a system are organized and under what conditions the parts of the system fail to operate. In brief, with mechanistic explanation one gains explanatory depth (Raerinne, 2013). Many ecological mechanisms are not well known (Raerinne, 2011). In fact, most mechanistic explanations in ecology are either underdetermined by the available data or by their absence (Raerinne, 2011, 2013). Thus, many causal explanations in ecology

are simple causal claims in the sense that there are no known or confirmed mechanistic explanations. When one contrasts ecology with genetics, molecular biology or neuroscience, disciplines where mechanistic explanations seem to be more prominent and better founded mechanistically, ecological causal explanations appear to be merely phenomenologically invariant generalizations (Raerinne, 2011). I would argue that the same might be said about the explanations provided by MTE.

In this sense, MTE's reference to allometries and scaling relationships should be understood as elucidating phenomena already known given available data in the sense of accommodating them. That is, allometries must be used to discover, describe, and classify the phenomena or *patterns* to be explained rather than being the things that do the explaining. Raerinne (2011: 261) illustrates this point with the following example.

Homeotherms (i.e., the organisms that exhibit the specific thermoregulation capacity of maintaining a stable internal body temperature regardless of external influence), poikilotherms (i.e., the organisms whose internal temperature varies considerably, being the opposite of a homeotherm) and unicellular organisms have different i_0 values (i.e., the normalizing constant independent of body size in the equations that relate metabolic rates to body mass in Eq. (3.2)). The values of i_0 are 4.1, 0.14 and 0.018 for, respectively, homeotherms, poikilotherms, and unicellular organisms. Some of the questions that need addressing concern the reason why the unicellular organisms have the lowest value of i_0 and how and why homeotherms metabolize at a higher rate (and therefore seem to use and exhaust relatively more resources) than poikilotherms and unicellular organisms of similar size. These are questions demanding an explanation. Knowledge of allometries spurs this kind of questions without, however, providing a direct answer; the only answer that is given is in terms of a phenomenological description, but surely *not* in terms of a mechanistic explanation. Nonetheless, allometries and scaling relationships may serve a heuristic role by suggesting new research questions, prompting the generation of new explanatory hypotheses and helping to discover regular connections between body size and other biological variables (Raerinne, 2013). Rather than providing explanations, many allometries and scaling relationships represent interesting objects of explanation, giving ecology interesting phenomena to be explained and the potential to progress and mature. This is a position close to that of some other authors, according to which ecologists seek regular patterns in nature even though they do not try to organize them within a body of theoretical explanations (Lawton, 1996). The reason is that they assume that ecology has no universal laws but just observable *tendencies* that cannot be derived from basic principles (Lawton, 1996; Colyvan, 2003). Accordingly, I would argue that most ecologists accept patterns of dependence as mere descriptions (as seems to me clearly the case with allometries and scalings), but *not* as causal explanatory relations. Instead, what they take to be explanatory is a description of the mechanism that produces these patterns. I would thus conclude that MTE cannot generally elucidate this mechanistic basis.

3.5 Conclusion

In summary, even though MTE has been praised for reinvigorating the study of metabolic and other allometries in ecology, thus aiming to capture interesting explanatory hypotheses, many criticisms have ensued. The theory aims to show how the metabolism of individual organisms affects the structure and dynamics of ecological systems, assuming that, at all levels, from individual organisms to ecosystems, the processing of energy and materials is linked through metabolic constraints and that the biogeochemical processes in ecosystems are largely consequences of the collective metabolic processes of the constituent organisms. Some ecologists (West et al., 1997; Enquist et al., 2003; Brown et al., 2004; Allen & Gillooly, 2007) think that metabolism is one of the great unifying processes in biology, making connections between all levels of organization possible, a theoretical hypothesis that I personally consider quite robust.

Notwithstanding this significant virtue, I would argue that the proposed (para) mechanistic hypotheses underlying MTE are problematic when tested, and that they are not fully consistent with fundamental aspects of ecology and physiology. My argument is underpinned by two critical considerations: that the cost of transport minimization is not a wholly tenable form of optimization by natural selection and that the Boltzmann normalization of thermal effects on metabolism requires a uniform response to temperature that sometimes is not observed. Due to these concerns, it is difficult to resist the conclusion that the assumed mechanisms underlying metabolic allometries postulated by MTE advocates are perhaps not the only ecologically relevant ones. From this point of view, the main problem of MTE is that the mechanisms postulated are treated as presumptively validated as well as mutually exclusive of other mechanisms proposed to explain allometries.

Thus, I find MTE lacking a complete explanatory mechanism for allometries and scaling relationships because: first, the proposed mechanisms are disconnected from the hypotheses they motivate; secondly, the putative mechanisms seem to be, biologically, not completely plausible; and, thirdly, the proposed universality of the mechanisms is hardly tenable. The theory, at best, highlights potential physical constraints imposed on the allometry of metabolic rates. It might therefore be considered as an oversimplified and insufficient description of the mechanisms underlying metabolic allometries. The upshot is that it cannot be considered a unifying model of mechanistic explanation in ecology for the simple reason that one must acknowledge that multiple mechanisms are likely to be involved in engendering the extant patterns of metabolic allometries. I would also argue that we cannot suppose that the so far illustrated criticisms are based, rather than on the analysis of available empirical data, on purely philosophical grounds. The “metaphysical” critics of MTE²⁷ seem to operate under the belief that unifying principles in biology and ecology, as

²⁷Those authors who base their criticisms not on data, but on deep-seated philosophical beliefs, using the expression of Allen and Gillooly (2007: 1076).

metabolism surely is, do not exist and that all species are unique. Consequently, they argue that the optimization of physiological traits is almost impossible in nature and that all aspects of natural selection are idiosyncratic and therefore unpredictable. I would argue that this latter critical perspective leaves no room for general predictive theories in ecology, and, for this reason, it is not a reasonable criticism of MTE. The MTE approach, in contrast, is based on the theoretical assumption that ecology is well served by the development of general quantitative theories yielding testable predictions, including how organisms will respond to environmental change. This theory is formulated on the premise that organisms share many common attributes, particularly with respect to metabolism, also assuming that there are general principles governing the process of evolution, and that these are inextricably linked to individual energetics, therefore embracing the principle of evolutionary optimization. I am not a “metaphysical” critic and I think that ecologists should not abandon the quest for unifying principles. The crucial issue is, as I have argued in Sects. 3.3 and 3.4, the oversimplifications adopted in the description of the putative mechanisms of the MTE, which become conspicuous when we consider the contradictory empirical evidence in support of MTE.

I suggest that we must recognize that a general theory such as MTE will never be able to explain *all* the variation of biological phenomena due to the inherent complexity of ecological systems and that, necessarily, further work is required, potentially testing all assumptions and predictions of its models. Nevertheless, I believe that MTE has succeeded in partially explaining a good range of phenomena at various levels of ecological organization and that it eventually holds some promise for somehow linking individual organisms to populations and then to communities and ecosystems using metabolism. Thus, MTE can turn out to be a valuable contribution to study some ecological phenomena within a mechanistic approach since it provides a deep study of *components* and their *activities* at various levels of ecological organization. The simple account of mechanisms in terms of delineated organized systems that operate in a linear start-to-finish sequence that is so prominent in the new mechanist literature is arguably a simplistic one as it does not often correspond to actual biological dynamics. My suggestion is that more integrated scale-free networks can point to a more accurate description of mechanisms as they operate in the actual world, while the appeal to free energy and its constrained release can also help to explain the nature of the activity of living mechanisms, a critical issue when we are linking the metabolism of individual organisms, populations, communities and ecosystems in a mechanistic model such as the MTE. In brief, my proposal is that we should not view mechanisms as entities in the world but as posits in mechanistic explanations that provide idealized accounts of what is in the world. It is reasonable, from this perspective, to argue that when scientists demarcate the boundaries of a mechanism, decompose it and localize its parts, they are following heuristic principles. As such, scientists use simplifying assumptions that facilitate the investigation, but which may turn out to be false, especially when they are dealing with systems that are non-sequential in nature and that are characterized by non-linear operations, such as all ecological systems are.

Acknowledgements This work was written with the support of the grant UI/BD/153731/2022 from FCT – Fundação para a Ciência e a Tecnologia, I.P. I would also like to thank the support of R&D Unity Centro de Filosofia das Ciências da Universidade de Lisboa (CFCUL), within the strategic project with the references FCT I.P. UIDB/00678/2020 and UIDP/00678/2020. I want also to express my gratitude to Davide Vecchi, for all his suggestions and requirements for the enhanced clarity of the text, and for showing me the recent new prospects of the New Mechanist Philosophy.

References

- Allen, A. P., & Gillooly, J. F. (2007). The mechanistic basis of the metabolic theory of ecology. *Oikos*, *116*, 1073–1077.
- Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge University Press.
- Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, *78*(4), 533–557.
- Bechtel, W. (2015). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Biological and Biomedical Sciences*, *53*, 84–93.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*, 421–441.
- Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *41*, 321–333.
- Bechtel, W., & Bollhagen, A. (2021). Active biological mechanisms: Transforming energy into motion in molecular motors. *Synthese*, *199*, 12705–12729.
- Bechtel, W., & Richardson, R. C. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research, 2010*. MIT Press.
- Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. (2004). Toward a metabolic theory of ecology. *Ecology*, *85*(7), 1771–1789.
- Colyvan, M. (2003). Laws of nature and laws of ecology. *Oikos*, *101*(3), 649–653.
- Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*, 547–563.
- Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across life sciences*. University of Chicago Press.
- Craver, C., & Tabery J. (2019). Mechanisms in science. *The Stanford encyclopedia of philosophy* (Summer 2019 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>
- Enquist, B. J., Economo, E. P., Huxman, T. E., Allen, A. P., Ignace, D. D., & Gillooly, J. F. (2003). Scaling metabolism from organisms to ecosystems. *Nature*, *423*, 639–642.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*, 49–71.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, *69*, 342–353.
- Glennan, S., Illari, P., & Weber, E. (2021). Six theses on mechanisms and mechanistic science. *Journal for the General Philosophy of Science*, *53*, 143–161.
- Hempel, C. H. (1965). *Aspects of scientific explanation: And other essays in the philosophy of science*. Free Press.
- Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across sciences. *European Journal of Philosophy of Science*, *2*, 119–135.
- Lawton, J. H. (1996). Patterns in ecology. *Oikos*, *75*(2), 145–147.
- Lawton, J. H. (1999). Are there general laws in ecology? *Oikos*, *84*(2), 177–192.

- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Nicholson, D. J. (2011). The concept of mechanism in biology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 152–163.
- O'Connor, M. P., Kemp, S. J., Agosta, S. J., Hansen, F., Sieg, A. E., Wallace, B. P., McNair, J. N., & Dunham, A. E. (2007). Reconsidering the mechanistic basis of the metabolic theory of ecology. *Oikos*, 116, 1058–1072.
- Pâslaru, V. (2009). Ecological explanation between manipulation and mechanism description. *Philosophy of Science*, 76, 821–837.
- Pâslaru, V. (2018). Mechanisms in ecology. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 348–361). Routledge.
- Pianka, E. R. (1966). Latitudinal gradients in species diversity: A review of concepts. *The American Naturalist*, 100(910), 33–46.
- Raerinne, J. (2011). Causal and mechanistic explanations in ecology. *Acta Biotheoretica*, 59, 251–271.
- Raerinne, J. (2013). Explanatory, predictive, heuristic roles of allometries and scaling relationships. *Bioscience*, 63(3), 191–198.
- Raerinne, J. (2017). Abstraction in ecology: Reductionism and holism as complementary heuristics. *European Journal for Philosophy of Science*. <https://doi.org/10.1007/s13194-017-0191-3>
- West, G. L., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276(122), 122–126.
- Wimsatt, W. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, 64, 372–384.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 4

Causing and Composing Evolution: Lessons from Evo-Devo Mechanisms



Cristina Villegas 

Abstract Evolutionary developmental biology (evo-devo) is often vindicated by theoreticians of the field as a mechanistic science that brings a mechanistic perspective into evolutionary biology. Usually, it is also portrayed as stressing the causal role that development plays in the evolutionary process. However, mechanistic studies in evo-devo typically refer to lineage-specific transformations and lack the generality that evolutionary explanations usually aim for. After reviewing the prospects and limits of a mechanistic view of evo-devo and their studies of homology and novelty, in this chapter I propose a way to combine the mechanistic view of evo-devo with the population-level inclination of more classical approaches to evolution. Such a proposal provides a philosophical framework for understanding the causal role of development in evolution both as mechanistic and as generalizable, population-level.

Keywords Evolutionary developmental biology · Innovation · Variation · Homology · Populations · Developmental repatterning

4.1 Introduction

If there is one type of explanation that has received the attention of most philosophers of biology in recent years, it is mechanistic explanation (Machamer et al., 2000; Glennan, 2002; Bechtel & Abrahamsen, 2013; Craver & Darden, 2013). So-called “new mechanicism” arose as a vindication of the non-nomological nature of many kinds of explanations in science, and it has been especially prolific in its application to biological phenomena. The mechanistic approach considers that there are scientific explanations without any appeal to fundamental laws of nature.

C. Villegas (✉)

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências,
Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal
e-mail: cvillegas@fc.ul.pt

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of
the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_4

61

These consist in describing a system that is responsible for the *explanandum* by decomposing it into its component parts and the activities such parts are engaged in, and subsequently recomposing the system back in terms of the way the parts and activities are organized to produce the phenomenon under study (Bechtel & Abrahamsen, 2013). Unlike more classical types of mechanicism, the new mechanical philosophy does not intend to *reduce* phenomena to the components of a system, rather stressing the role of the organization of those component parts into a whole in bringing about phenomena. This feature makes mechanicism especially interesting for the life sciences.

Mechanistic explanations are particularly prominent to account for what Mayr once (1961) labeled *proximate* causes, namely the causes acting at the level of an organism, determining how it is and what it does. Unlike *ultimate* or evolutionary causes, proximate causes are responsible for the functioning and behavior of organisms. If a scientist seeks to explain *how*, say, hearts pump blood, then she needs to refer to the mechanism responsible for such pumping, which may include factors such as muscle contractions, regardless of the evolutionary history of the circulatory system. A well-known example is the mechanism for the circadian cycle, introduced by Bechtel and Abrahamsen (2010) to exemplify the features of mechanistic explanations. In this example, the relevant components are biochemical substances within cells, such as RNA, which act by regulating the presence of further biochemical components, such as proteins.

Evolutionary developmental biology (evo-devo for short) is often vindicated by theoreticians of the field as a mechanistic science (Wagner et al., 2000; Hall, 2003; Müller, 2007). This mechanistic aspect is important insofar as evo-devo is considered to bring a mechanistic perspective into the otherwise “mechanism-less” field of evolutionary biology. In particular, it is argued that taking development into account entails considering the mechanisms underlying phenotypic change, a precondition for evolution through natural selection that is assumed but not explained in classical approaches to evolution. From a classical perspective, evolutionary questions pertain to a separate domain of biological causes altogether than mechanistic ones: ultimate causes acting on populations throughout generations (Mayr, 1961). Some philosophers have argued that, in fact, evo-devo shows that there is no such separation between ultimate and proximate causes, and that, instead, there is a reciprocal causation between organismal and evolutionary causes (Laland et al., 2011). Nonetheless, combining these different kinds of explanation, “ultimate” or evolutionary, and “proximate” or mechanistic, is germane to this field. As Ron Amundson holds, the “difficulty of integrating population thinking with the mechanistic thinking of developmental biology” is inherent to evo-devo (Amundson, 2015, p. vii).

This scenario gets more complicated when further subtleties about biological causation are introduced. A more complete, still classical picture of biological causes is Tinbergen’s (1963) categorization based on the four questions to be asked about the nature of traits: their survival value, their evolutionary history, their ontogenetic origin, and how they work in a *mechanistic* sense. Setting aside some ambiguities (see Conley, 2020), Tinbergen’s schema is typically interpreted as specifying two distinct types of *proximate* causes *à la* Mayr: mechanistic explanations and

explanations of ontogeny (Bateson & Laland, 2013). Here, “mechanistic explanations” focus on physiological aspects independently of the way the traits are acquired during ontogeny, whereby this acquisition process represents a different type of proximate causal process. The most classical works on biological mechanisms seem to prove this association right, insofar as most mechanisms refer to the way some system works rather than how it came to be acquired during development, as in the classical biochemical example of the circadian cycle (Bechtel & Abrahamsen, 2010).

Paying attention to this schema, the abnormal situation of evo-devo becomes clear. Not only does it combine the proximate and the ultimate domain by bringing together organismal and evolutionary causes, but it does so through the consideration of *ontogenetic* causes rather than *physiological* ones. Some philosophers have pointed out the difficulty of applying the mechanistic schema to ontogeny, and thus to the domain of developmental biology (Mc Manus, 2012; Love, 2018; Baedke, 2021), which further complicates the task of applying it to evo-devo. Importantly, this difficulty concerns the causal nature of developmental mechanisms, and raises the question of whether developmental mechanisms are a cause of evolution, as argued by evo-devo theoreticians (Wagner et al., 2000; Müller, 2007). In the upswing of the new mechanistic philosophy, and given the apparent centrality of mechanisms in evo-devo (Baedke, 2021), it becomes important to analyze to what extent evo-devo is a mechanistic science and, moreover, what that says about the causal relation that development holds with evolution. That is the aim of this chapter.

The chapter is structured as follows. Section 4.2 reviews the prospects and limitations of considering evolutionary and developmental biology, separately, as mechanistic sciences. Section 4.3 addresses the mechanistic aspects of evo-devo studies of homology, arguing that they point at a causal and compositional¹ role of developmental mechanisms in phenotypic unity and diversity. Section 4.4 discusses the causal role of development in phenotypic changes, both in evo-devo studies of evolutionary novelties and within the broader domain of a mechanistic view of evolution. It introduces the idea of *developmental repatterning* (Arthur, 2011) as a population-level evolutionary mechanism responsible for biases in phenotypic change and innovations. Section 4.5 concludes with some final remarks about causation and mechanicism in evo-devo.

4.2 Two Unusual Kinds of Biological Mechanisms

Evo-devo is a highly interdisciplinary research area that combines insights and methodologies from developmental and evolutionary biology, both broadly construed. It brings a comparative and phylogenetic perspective to the study of developmental systems and, moreover, it sheds light into the developmental

¹In this chapter, I use constitution and composition interchangeably.

processes underlying phenotypic change in evolution (Müller, 2007). Before assessing the mechanistic and causal aspects of evo-devo as a discipline, it is therefore convenient to review the scope and limitations of mechanicism for each of the disciplines that it intends to combine.

4.2.1 *Mechanisms of Development*

Developmental biology studies the process of organismal formation from zygote to adulthood (or some other developmental stage). Every multicellular organism goes through this process, which consists of a myriad of physical, molecular, cellular, tissue-level, and organism-level sub-processes. Despite the ubiquity of development within multicellular life, developmental biology largely lacks general laws (Minelli & Pradeu, 2014), instead focusing on species or taxa-specific processes, typically through the study of model organisms such as the fruit fly *Drosophila* or the weed *Arabidopsis*. Like many other branches of the life sciences, developmental biology was originally descriptive, providing phenomenological models of embryonic stages with little to no mechanistic content. But innumerable progresses have made developmental biology grow as an experimental science that identifies relevant components and causal factors in the formation of specific traits in organisms, making it a good candidate for being a field where mechanisms play an important explanatory role. For example, in the era of developmental genetics, there is a great deal of developmental biology that relies on experimental studies of correlative associations of genes. For instance, knockout experiments identify genes that *make a difference* in the adult phenotype by blocking their expression at a given developmental stage. Building mechanistic models of how such a difference is made is a crucial part of how developmental biology seeks to explain.

Let me now introduce an example that will guide us throughout the chapter. Flowers are the reproductive traits of angiosperm plants, a monophyletic group that separated from their closest extant relatives, gymnosperms (plants that generate unenclosed seeds, such as conifers), between 300 and 350 million years ago. They are not individual organs but a highly integrated set of them. While there is much flower diversity, a prototypic flower usually consists of at least four organs: sepals, petals, stamens, and carpels. The development of flower organs follows a pattern known as “the ABC model”, which associates a specific combination of three to five homeotic genes (labeled *A*, *B*, and *C*; or *A*, *B*, *C*, *D* and *E* in more recent versions) to each of them (Theißen & Rümpler, 2021). Although only more complex versions of it seem to be currently accepted (Theißen et al., 2016), the classical model can serve for illustrative purposes. The basic idea of the model is that *A*, *B*, and *C* genes are expressed differentially in specific developmental stages, forming flower organs in a temporal and spatial sequence. For example, the original model states that expression of *A* genes forms sepals, while the combination of *A* with *B* genes forms petals. The sequential expression of, first, *A* genes and, then, *B* genes would then make sepals appear before and below petals in flowers. The model is not

mechanistic *per se*; rather, it identifies genes as *difference-making* causes for specific phenotypes through mutational experiments. However, mechanistic versions of the model identify the transcription factors that ABC genes code for, as well as the proteins that their target genes produce. It is the case for the “floral quartet” model (Theißen et al., 2016), where MIKC-type proteins derived from ABC genes form quaternary complexes that bind to specific DNA regions to control the expression of genes involved in the formation of the different flower organs. With this example in mind, let us see how to make sense out of developmental mechanisms.

The extent to which there is a defined set of entities that can constitute a developmental mechanism is not a simple question, but much work has been devoted to uncovering these entities in the last decades. Gene products within cells such as proteins and transcription factors are the main components of development at the chemical level, for they are responsible for important developmental factors such as signaling and gene-regulatory pathways that translate environmental and genetic inputs into phenotypic outputs such as cell differentiation. Extracellular and environmental components can also be crucial, as well as physical components such as mechanical stresses or bioelectrical potentials, all of these responsible for the behavior and expression patterns of cells at specific points of the developing organism. The number and kinds of components that are relevant for explaining a specific developmental phenomenon not only are vast but vary for each specific case. Still, the nature of these components is relatively well understood in model systems. In the model of flower organ development illustrated above, the entities specified are MIKC-type proteins that ABC genes code for, the sections of DNA where these proteins bind, and the gene products of those sections.

The complications arise when it comes to understanding the organization of activities carried out by developmental mechanisms (Jaeger & Sharpe, 2014). Developmental processes are dynamically complex. Explanations of developmental mechanisms shall count as *dynamic* mechanistic explanations, that is, as those in which the recomposition of the system involves the description of a nonlinear dynamic behavior (Bechtel & Abrahamsen, 2010). The regulation of specific variables such as gene expression is normally non-linear and allows, for example, for genes regulating their own expression, so that the variables and parameters of developmental systems tend to be time-dependent (Jaeger & Sharpe, 2014), turning the whole parametrization of the dynamics into a complex task. The result is complex behaviors such as threshold effects and feedback loops. For instance, in the ABC model of flower development, some protein complexes bind to two gene regions forming DNA loops (Wagner, 2014). In these cases, gene expression only results from a complex pattern of cooperativity and interdependency within the protein complexes and with their binding sites.

Some philosophers have pointed out that these complex developmental dynamics uncover some limitations of the mechanistic paradigm (Mc Manus, 2012; Love, 2018; Baedke, 2021). In explanations of non-developmental phenomena such as cell metabolism, the *explanandum* is extended in time but remains constant. Developmental phenomena, in contrast, are not only extended in time but of changing nature. When developmentalists seek to explain the development of a trait, such

as petals and sepals in flower plants, they need to account for profound changes in the nature and organization of its mechanistic elements, including in their hierarchical organization in levels. This is because the organ level, where we may situate *the flower*, is only the result of the process, but is not present in all the stages that precede its formation. For example, the floral meristem is a set of undifferentiated cells that precedes the flower but is not yet identifiable as a differentiated organ. Additionally, the flower forms out of the expression in meristem cells of E-class genes, which are a molecular rather than a physiological reality (unlike the floral meristem).

The conundrum of this problem relates to the causation/constitution distinction. When first introduced, the new mechanistic philosophy acknowledged a dual potential of the framework: in terms of *causal* explanations and in terms of *constitutive* ones (Craver, 2007; Bechtel & Abrahamsen, 2013; Romero, 2015). Causal explanations are those where *explanans* and *explanandum* are cause and effect, which demands that they are not the same entity and that there is an asymmetric temporal relationship between one another. A gene involved in the formation of the floral meristem can be a cause of the proteins responsible for meristem cell differentiation. In contrast, constitutive explanations are those where *explanans* and *explanandum* are different levels of description of the same system, which implies their ontological identity and temporal synchronization. A petal cell can be a constitutive part of the flower. Thus, a mechanistic explanation of a flower can mean two different things: a description of a *causal* mechanistic chain of events at the morphological level leading to the formation of the flower, or a description of a *constitutive* mechanism at a lower level such as the molecular. Developmental explanations, nonetheless, typically combine both (Mc Manus, 2012; Ylikoski, 2013). For example, cellular properties in the flower meristem can be both seen as causal factors and lower-level components in the formation of flower organs.

Given this complexity, other frameworks have been proposed for understanding both the nature of development and of developmental explanations. The processual view of development (Baedke & Mc Manus, 2018; Nuño de la Rosa, 2018; Bich & Skillings, 2023) argues that developmental processes are temporal parts of the life cycle, where dynamic organization plays a more fundamental role than the entities it gives rise to. For example, Baedke and Mc Manus (2018) contend that a better way of understanding hierarchical levels in development is to consider *time scales* instead of compositional relations. Similar arguments are found in the dispositional view of development (Hüttemann & Kaiser, 2018), where mechanisms are seen as the manifestation process of the dispositions of a system. By invoking a primordial dynamic developmental reality from which mechanisms can be abstracted away in scientific practice, these two positions avoid some of the complexities of understanding mechanisms as the fundamental ontology of development.

Nonetheless, and despite the complications derived from adjusting the complex science of developmental biology to the mechanistic framework, it is beyond doubt that developmental biologists explain by describing the components of developmental processes and their organizational properties, including their causal, spatial, and also temporal relations. The cruciality of these temporal relations, while

challenging, need not undermine a mechanistic framework, especially given that the new mechanistic philosophy constantly embraces new ways of accounting for complex dynamics (see Krickel this volume). For example, it has been recently argued that the distinction between causal and constitutive relations should be abandoned altogether since, at least in some sciences—including biology—, “diachronic causal constitutive relations” are instead the norm (Leuridan & Lodewyckx, 2021). According to this argument, the constitutive relations described in science are sometimes diachronic, implying some level of temporal extension where interlevel causation also plays a role. Such a consideration might help accommodate the fact that new levels of mechanistic organization are generated in development (e.g., the emergence in developmental time of the tissue level from pre-existing lower-level elements such as cells), a phenomenon that is sometimes considered to be out of the scope of mechanistic explanations (Baedke & Mc Manus, 2018; but see Austin, 2016 for a different view). In sum, whether or not there are currently enough philosophical tools to integrate developmental complexity, it seems that a mechanistic view has a strong potential to incorporate most of the developmental phenomena. How exactly to do so—and whether this view must be ontological or epistemological—remains an open question, which additionally complicates the task of accommodating these unusual kinds of biological mechanisms into the evolutionary domain.

4.2.2 *Mechanisms of Evolution*

Evolution is the historical process of transformation and diversification of the tree of life. The process itself is thought to follow similar rules of descent with modification throughout the whole tree, and finding general models on the basis of such rules is one classical target of evolutionary biology. One of its goals is thus to explain the phenotypic composition of species in terms of the rules governing the historical changes undergone by populations—and sometimes to predict short-term future ones. Unlike developmental biology explanations, these explanations are not usually mechanistic, in the sense that they don’t consist of a decomposition (and recomposition) of evolutionary phenomena into entities and activities. Rather, they typically take the form of the application of statistical models involving population dynamics factors such as natural selection, genetic drift, and mutational rates.

As mentioned in the introduction, Mayr’s (1961) classical picture labels these factors *ultimate causes*. However, some philosophers reject the idea that one can talk about causes at the level of population dynamics (Walsh et al., 2017, reviewed in Pence, 2021). Their viewpoint is that the classical population-level explanations of evolution do not reflect evolutionary causes, because the real causes of evolution act at the level of individuals and their relation to the environment (Walsh et al., 2017). From this position, “ultimate” explanatory terms such as genetic drift are statistical, and only organism-level explanations of the trends they represent provide causal explanations of evolution. This alleged lack of causal content in classical evolutionary explanations is sometimes referred to as a lack of *mechanistic* content

(Pigliucci & Kaplan, 2010). From this point of view, organism-level mechanisms are considered causally responsible for the changes that underlie the population-level trends we see in evolution. One task for the agendas extending the classical framework of evolution, including evo-devo, is therefore to study the interplay between those individual-level mechanisms and their population-level effects (Laland et al., 2011). For example, evo-devo studies how differences in the mechanisms of flower organ development may have resulted in patterns of selection (Mondragón-Palomino et al., 2009).

Other scholars defend that the causes of evolution are fairly described at the population level. Organism-level mechanisms such as developmental or ecological ones *compose* evolutionary causes at a lower level from this point of view. The rationale is that, since they are not difference makers of populational changes, these organism-level developmental and ecological mechanisms need not be considered in causal evolutionary explanations (Millstein, 2003). There are different ways of interpreting the population-level causes present in classical evolutionary explanations, including the mechanistic perspective. Thus, a number of philosophers have provided tentative analyses of what a mechanistic view of population-level evolutionary causes would look like (Skipper & Millstein, 2005; Barros, 2008; Illari & Williamson, 2010; DesAutels, 2016, 2018). From this position, organism-level developmental and ecological mechanisms can be considered components of the population-level evolutionary *mechanisms* referred to in models of population dynamics. One mechanistic task for evo-devo would therefore consist in specifying the role of development in population-level mechanisms of evolution.

The philosophical enterprise has nonetheless mostly been limited to natural selection, following the usual label in the scientific literature, where selection is often called a “mechanism” of evolution (e.g., Bell, 2008). In addition, most of the attempts have focused on pointing out limitations of the first new mechanistic views to account for selection, and considerations of more recent mechanistic developments are rare. For example, Skipper and Millstein (2005) first noticed that the irregularity of evolutionary phenomena made it unsuitable to be analyzed in terms of the early mechanistic views proposed by Machamer et al. (2000) and by Glennan (2002), and suggested that only a “stochastic” or probabilistic mechanistic approach could deal with such difficulties. However, while most later mechanistic views acknowledge the stochastic nature of mechanisms, the irregularity of natural selection may not entirely be captured by stochasticity. As Pérez-González and Luque (2019) point out, this irregularity is germane to the fact that natural selection always acts in conjunction with other evolutionary factors such as mutations and migrations, and is inseparable from its counterpart genetic drift. A population that does not change at the rate predicted by selection is, by definition, a population that undergoes drift. This irregularity problem applies more generally to any attempt to understand selection in mechanistic terms insofar as it is hard to identify the specific phenomenon that it produces as distinct from the results of other evolutionary factors (Beatty, 1984). This is why functional approaches identify the phenomenon that selection explains with a specific *process* rather than an outcome (DesAutels, 2016),

pointing at the events underlying the higher survival and reproductive success of fitter organisms in populations.

Similarly to the worries raised about dynamism for developmental mechanisms, the foundational article from Skipper and Millstein (2005) further argued that selection is better described as composed of stages or time-slices rather than parts. In addition, it stressed that the interactions that comprise it do not fit the standard criteria of a mechanistic view, since the relevant activities attributed to selection (such as those organisms engage in during reproduction and in their relation to the environment) are at the very least suspicious of lacking regularity. One attempt to solve this consists in relaxing the criteria of stability of both entities and activities. According to Illari and Williamson's (2010) view, these need only be *functionally stable*, that is, stable enough to produce the phenomenon of natural selection. Following their account, selection fits into a "functional hierarchy" composed of a myriad of mechanisms acting at different lower levels to produce the enhanced survival and reproduction of fitter individuals. These lower-level mechanisms are composed of very diverse types of entities such as populations, organisms or chromosomes, and activities like sexual selection, recombination, or reproduction, presenting organization insofar as they combine to produce selection.

It is in this sense that developmental and ecological mechanisms are perceived as components of population-level mechanisms, although there is no consensus about how to understand this compositional relation. For example, while Illari and Williamson (2010) advocate for selection as a highly complex multilevel mechanism, authors such as Barros (2008) stress that selection is a two-level mechanism: it acts at the level of the individual-environment interaction and at the population level.

However, there is no point in discussing the components of selection without a clearer view of how they relate to other causes of evolution, such as drift or mutations (DesAutels, 2018; Pérez-González & Luque, 2019). In particular, functionally individuating the lower-level mechanisms composing selection demands criteria for discerning when they compose other higher-level causal factors of evolution. Indeed, disagreements about the levels composing selection as a mechanism might indicate a disparity in the evolutionary phenomenon it is supposed to explain. Here is where non-classical agendas of evolution may be of help, since describing the complexity of evolutionary mechanisms and how they interact is part of their agendas. In the case of evo-devo, one goal is to understand how evolutionary mechanisms relate to developmental ones. The issue actually comprises two different aspects of evolutionary mechanisms. The first one is: are other population-level causes of evolution mechanisms too? As a matter of fact, very few attempts have been made to account for genetic drift, mutations or migrations in population-level mechanistic terms (DesAutels, 2018), and there seems to be no framework for understanding the way in which these putative mechanisms may interact with one another (Pérez-González & Luque, 2019). The second one is: are the lower-level mechanisms composing evolution *evolutionary causes* too? The mechanistic view of evolution seems to involve causes and mechanisms at different levels and, as such, needs to deal with how these different levels of causation relate to one another. With these questions in mind, let us turn to the mechanistic aspects of evo-devo.

4.3 *Evo and Devo: The Mechanistic Composition of Variation*

The main explanatory agendas where evo-devo clearly introduces a mechanistic perspective are the study of homology and of evolutionary novelty. These are the two sides of the same coin: novelty is present whenever a trait lacks an homologue in its ancestral lineage and is not homologous to a different body part of the same organism. In these studies, developmental mechanisms are seen in a comparative and phylogenetic framework, and the goal is to explain the mechanistic bases of phenotypic commonalities and diversity. This is an important task in the evo-devo agenda, for it brings a mechanistic aspect to the study of phenotypic variation, which is a necessary condition for evolution to take place, but mostly assumed without explanation in classical evolutionary approaches.

Intraspecific variation—the one that matters for classical, microevolutionary approaches—implies the existence of the same trait in different forms in a given population: for instance, bigger or smaller versions of the same wing in a bird species, brighter or darker color versions of the same petal in a plant species, etc. This phenotypic variation always implies variation in the operation of the same developmental mechanisms, be it a slightly different interaction among its components, a different concentration of some of its elements, or a different interaction between the mechanisms and their environment. But when one looks at the inter-specific level, the issue gets more interesting. Homology is the presence of the same unit (an organ, a cell type, a morphological trait, a behavior) in different species. The study of homology is central to evo-devo because it tells us about what has been preserved in evolution and how it varies. One central question for evo-devo is thus whether there is homology of the developmental mechanisms responsible for homologous traits. In other words, whether there is correspondence in homology across different levels of composition of phenotypic diversity. This question gets right at the problem of characterizing developmental mechanisms, for it concerns how much variation in the developmental mechanism corresponds to variation in the phenomenon it explains, i.e., the phenotype.

It has been argued that developmental mechanisms have a “hybrid nature” (Newman, 2014): molecular and physico-cellular. While much of developmental biology focuses on the molecular level, physico-cellular mechanisms can be described for most developmental processes, complementing the description in terms of molecular ones. For example, in the case of flower development, cell division in the floral meristem prior to the activation of ABC genes depends on the radial position of cells, a physical property, regardless of the specific gene expression profile of the cell in question (Alvarez-Buylla et al., 2010). This could at first sight simply be interpreted as concerning the nested nature of mechanisms, where molecular genetics mechanisms (lower-level) compose cellular-physical ones (higher-level), but it actually concerns a much deeper developmental problem (Love, 2018). Comparative studies at different levels of organization have made it plain that seemingly homologous traits may be realized by a multitude of

mechanisms at a lower level, while apparently diverse traits can be realized by similar developmental mechanisms (DiFrisco et al., 2020). Thus, many different genetic mechanisms are constitutive of the same higher-level, morphogenetic mechanism in different species. And the opposite also holds: many morphogenetic mechanisms of development, such as tissue-level mechanical forces, vary more across species than the molecular mechanisms supposedly composing (and causing) them.

This poses a serious challenge for evo-devo research: are conserved mechanisms responsible for the generation of homologous traits despite this pervasive diversity? Or is homology of traits acquired in other ways relatively independent of genetic and developmental conservation? The two positions illustrated by this (even if very simplified) dichotomy indeed represent a significant divide in approaches to the problem of homology in evo-devo (Nuño de la Rosa & Etxeberria, 2012). On the one hand, character identity views of homology postulate the high conservation of core developmental mechanisms that provide “identity” to traits inasmuch as they are involved in the production of variants of the same character under an array of developmental contexts (Wagner, 2014; DiFrisco et al., 2020, 2023). On the other hand, organizational views of homology hypothesize positions of phenotypic stability where different developmental mechanisms converge in virtue of the internal organization of body plans, and that it is this convergence what characterizes trait homology, rather than the sameness of mechanisms underlying phenotypes (Müller, 2003; Newman, 2003; Peterson & Müller, 2016). The idea of mechanism is central to both positions, but for different reasons. Let me explain.

The “character identity mechanisms” view of homology proposes that there are level-specific mechanisms that explain homologous traits, and are responsible for their traceable identity in evolution at different levels (DiFrisco et al., 2020, 2023). Character identity mechanisms are “mechanistic architectures” with specific causal profiles that are retained in evolution despite changes in the inputs activating them as well as in the realization of their phenotypic effects:

[Character identity mechanisms] are less replaceable in evolution than their upstream signaling inputs and downstream effector mechanisms ... [A]s a result, ... [they] are more likely to be evolutionarily conserved than other developmental mechanisms. (DiFrisco et al., 2020: 8–9).

The key idea is that the persistence of traits in evolution is explained by the recurrent instantiation of mechanisms causing and composing them in virtue of their parts, activities and organization. As mentioned, the nature of these mechanisms can differ depending on the level of organization of each phenotypic trait. For example, for cell types, a character identity mechanism is postulated to be composed of a gene regulatory network with cross-regulatory and signaling activity, while for tissues it is composed of cell types, extracellular matrices, signaling molecules, and the inter-cell signaling complexes they all engage in (DiFrisco et al., 2020). Depending on which specific cell type or tissue identity the mechanism is responsible for, the specific nature of these entities and activities changes. In addition, while these mechanistic profiles remain stable in different species, the overall process they are part of may undergo several types of changes. For example, they

specify the identity of the trait but not its “character state” or specific realization (Wagner, 2014; DiFrisco et al., 2020). The ABC model of flower development could exemplify this idea, for it describes a pattern found across all angiosperms despite the overwhelming diversity in flower morphology (e.g., size, shape, color, number and arrangement of each of the organ types). According to this view, this points at the retention of the elements and organizational lower-level mechanistic structure responsible for the identity of flower organs, like ABC genes, but not of other mechanistic components that explain the specific features of particular, species-specific flowers. Interestingly, this retention is not always explicable in terms of natural selection, for the causal profile of character identity mechanisms is sometimes independent of function. For example, most flowers are bisexual, with both stamens and carpels—male and female organs, respectively. However, in the rare cases of unisexuality in flowers, stamens and carpels develop too, only that one of these organs is sterile. This suggests that the floral plan, which includes sepals, petals, stamens and carpels, and is explained by the ABC model of development, may be developmentally retained despite changes in function, and thus that selection is not the only responsible for the co-occurrence of floral organs (Wagner, 2014).

From the perspective of the organizational view of homology (Müller, 2003; Newman, 2003; Peterson & Müller, 2016), traits are retained because of their organizational role in development and inheritance. However, unlike in the previous approach, traits can be homologous independently of their mechanistic composition at any given level. That is, different developmental mechanisms can converge to produce homologous traits. Also, two lineages may retain a homologous trait even if the underlying mechanisms for it undergo severe, independent changes. Therefore, in this case, the hypothesis is that variation in developmental mechanisms does not correspond entirely to variation in phenotypes. From this view, one may argue that, instead, traits are individualized in virtue of their causal and compositional *role within the entire body plan*, the latter understood in mechanistic terms or not. That is, while the identity of a trait is independent from its mechanistic basis, it depends on its specific organizational role within the developing organism and in reproduction:

Homology denotes constancy of constructional organization despite changes in underlying generative mechanisms ... Homologues act as organizers of the phenotype ... [and] as organizers of the evolving molecular and genetic circuitry (Müller, 2003: 64).

If we understand the organism and reproduction in mechanistic terms, then traits are individualized as specific mechanistic components engaged in specific constitutive and causal activities. I will use the ABC model of flower development for exemplifying this idea too. From the organizational view of homology, the developmental pattern found across angiosperms does not point at the retention of any specific lower-level mechanistic element (such as a particular protein type). Rather, it points at the organizational role of the pattern itself within the development of angiosperms. As in the case with the character identity view of homology, the retention of the pattern does not depend solely on natural selection. In this case, it depends on the whole developmental organization of flowering plants.

These two different evo-devo approaches to homology provide mechanistic pictures of what prevails and varies in evolution, partially explaining patterns of interspecies phenotypic variation. Taking the perspective of developmental mechanisms as both causal and constitutional (Ylikoski, 2013), they both give a causal and constitutive partial explanation of extant variation. On the one hand, character identity mechanisms (DiFrisco et al., 2020) stress what are the relevant causal and compositional *lower-level relations* that make a developmental mechanism instantiate a specific type. On the other hand, the organizational view of homology (Müller, 2003) emphasizes what are the relevant causal and compositional *organismic-level relations* that make a developmental system generate a specific phenotype.

4.4 Causing Phenotypic Change

As mentioned above, homology and novelty are the two faces of the same coin. Thus, having a criteria for what counts as homology is tantamount to having criteria for discerning what is an evolutionary novelty. However, evo-devo is not just concerned with what counts as a new trait. It also seeks for mechanistic explanations of the phenotypic changes raising novelties.

Traditionally, phenotypic change has been associated with factors external to the organism, such as the occurrence of mutations or recombination through directional selection. From an evo-devo perspective, however, what matters is that external factors trigger phenotypic changes that significantly depend on the properties of the developmental system. In the case of evolutionary novelties, external factors can be seen as “initiating conditions” of phenotypic change, while developmental systems act as the “realizing conditions” of those changes (Müller & Newman, 2005). Taking a mechanistic approach to evo-devo, it follows that the features of developmental mechanisms causally contribute to the directionality of evolutionary change. In this last section, I revise some mechanistic aspects of evo-devo studies of evolutionary novelty (Sect. 4.4.1), and I introduce the idea of developmental repatterning as a population-level mechanistic component of evolution (Sect. 4.4.2).

4.4.1 Mechanistic Views of Innovation

Evo-devo approaches to homology also postulate mechanistic views of how evolutionary novelties may arise. From the “character identity mechanisms” perspective, evolutionary novelties can be explained by a reuse (through co-option) of a mechanistic component of one trait into another identity mechanism (DiFrisco et al., 2023). Some genetic changes lead to the reuse of the same developmental components, often by duplication, into a different trait. This reconfiguration of the mechanism underlying the trait may give rise to an evolutionary novelty. On the other hand, the organizational view of homology sees the very process of innovation as a

mechanism based on developmental properties (Peterson & Müller, 2016). Here, innovation is depicted as a number of stages where new components and activities arise from the previous ones, giving rise to new homologues (Müller, 2003; Newman, 2003). The idea is that the origin of traits is mostly driven by the cell, tissue, and epigenetic-level processes in a first phase of generation, before being genetically accommodated through canalization or similar processes at a phase of integration. Finally, traits become increasingly independent from the mechanisms of innovation involved in their generation (Müller, 2003).

However, providing mechanistic details of how novelties arise requires a complex picture that is rarely attainable. Philosopher Brett Calcott pointed out that most explanations of novelty in evo-devo take the form of lineage explanations, which actually consist of series of independent mechanistic explanations (Calcott, 2009). An evo-devo lineage explanation gives a set of mechanistic models of a developmental system, each of which explains (constitutes and causes) a particular stage in an evolutionary series of phenotypic variants. The requirement for such a set to constitute a lineage explanation is that each mechanistic model is linked to the next one by a continuity requirement. This means that one mechanism must be similar enough to the next one so as to justify that the one could be the result of a minor modification in the other. Therefore, one shall point at the right components in decomposing a developmental mechanism, because it is continuity in these components that warrants the plausibility of lineage explanations (Calcott, 2009).

It is important to stress that lineage explanations don't provide mechanistic explanations of evolutionary change, but a series of plausible (gradual) evolutionary changes in mechanisms (Kaiser, 2021). For the origin of flowers, evo-devo models have proposed that small changes in gene expression, such as the accumulation of higher levels of protein concentration or the acquisition of new binding sites, could have gradually resulted in the emergence of the ABC pattern of flower development (Baum & Hileman, 2006). But the causes of such changes are typically left outside the lineage explanation (Calcott, 2009): what caused the incremental changes in protein concentration or binding sites number is external to the developmental mechanisms described, and therefore not included in the explanation of the origin of the developmental pattern. Additionally, many evolutionary novelties involve the emergence of new levels of mechanistic organization, meaning that continuity between mechanisms does not always imply graduality.

Following the continuity requirement involves knowledge about *potential changes* in developmental systems. Several authors have pointed out that individuating developmental systems indeed demands a consideration of this potential, understood in dispositional (Brigandt, 2015; Austin, 2017) or topological (Jaeger & Sharpe, 2014) terms, thus falling beyond the limits of what a mechanistic account can deliver. This is indeed a crucial aspect of the evo-devo research agenda: it combines mechanism-based knowledge with quantitative means of explanation in order to introduce development into the broader evolutionary domain (Brigandt, 2015). Here, evo-devo typically abstracts away from specificities of developmental mechanisms, often making use of dispositional or even mathematical means of explanation. Hybrid explanations, where dispositional and topological explanations are

combined with mechanistic insights, are actually the norm in most evo-devo cases (Brigandt, 2015; Huneman, 2018).

This is not just a methodological observation. In the context of evo-devo, the issue of what a developmental mechanism is remains naturally intertwined with the issue of what kind of phenotypic transformations it can undergo in evolution, which normally demands explanatory means beyond the decomposition and recomposition of a system. For instance, Jaeger and Sharpe (2014) point out that developmental mechanisms shall be identified by a particular way of bringing about phenotypic changes, based on topological similarity in a configuration space of possible phenotypes. Similarly, Austin (2017) stresses that what characterizes an evo-devo ontology is that:

it is [developmental] systems' intrinsic generative capacities which are causally responsible for providing the morphological novelty which subsequently shapes the *evolutionary* (read: selective) *landscape*. (Austin, 2017, p. 377, stress added).

In other words, the internal properties of developmental systems are responsible for their own variational tendencies in evolution, i.e., for the way they can vary (Wagner, 2014). This is why the bridge between developmental insights and the evolutionary approach has been vindicated in terms of the *variational dispositions* (Austin & Nuño de la Rosa, 2021) or *propensities* (Nuño de la Rosa & Villegas, 2022) that a developmental system has insofar as it has a tendency to generate evolutionary variation in specific ways. Mathematical means for measuring these propensities are indispensable, as exemplified in the use of the genotype-phenotype map as a mathematical instrument for predicting phenotypic changes from genotypic ones, and the use of morphospaces for studying the feasibility of evolutionary transformations. These methodologies typically intend to bridge evo-devo approaches to population level studies of evolution. Studies in flower evolution also exemplify this evo-devo approach, where the use of floral morphospaces, or mathematical spaces of possible flower phenotypes, are tools for studying the evolutionary dynamics of flowers in terms of both selective and developmental factors (Chartier et al., 2014). Although introducing mechanistic knowledge into genotype-phenotype mapping is an emergent tendency for increasing their predictive accuracy (Pavličev et al., 2023), there is a trade-off between this accuracy and their generality. In sum, it seems that evo-devo studies of novelty and developmental innovation are not always improved by increasing the level of mechanistic content (Brigandt, 2015).

4.4.2 *Innovation as an Evolutionary Mechanism*

We are now left with the task of relating evo-devo explanations to a mechanistic picture of evolution. The previous sections summarized the mechanistic view that is mostly regarded in evo-devo: an analysis of how specific mechanisms of development vary or can vary in evolution. This section concerns another interest of evo-devo, namely how development *systematically* biases the production of evolutionarily

relevant variation in lineages and populations. Notice the difference here. We have seen that developmental mechanisms are involved in the causation and composition of the phenotypic variation available for other evolutionary factors to feed upon. What we have not seen yet is whether there is a way to integrate the role of developmental biases in population-level causal explanations. For this, we need to assess how developmental mechanisms relate to populational causes, including alleged evolutionary mechanisms such as selection, drift and mutations.

One way to do this is to adhere to the statisticalist view of evolution (Walsh et al., 2017), and consider that developmental mechanisms, acting at the level of organisms, are causally involved in the only process that matters to evolution: the life cycles of individuals. These mechanisms make a difference to the way the reproduction of organisms give rise to new phenotypes. In particular, a mechanistic understanding of development contributes to a finer-grained picture of evolution by providing a mechanistic (partial) description of the generation of new variants in a specific lineage. Mutational and recombination events would trigger the response of developmental mechanisms, which constrain the phenotypic outcomes those triggers can generate (Baetu, 2012). In this picture, there are no ultimate causes: there are only *proximate* causes, in this case *ontogenetic*, that engage in a relation of reciprocal causation with the environmental needs of organisms (Laland et al., 2011).² This process of reciprocal causation *explains* the statistical trends taking place at the population level.

One could be satisfied with this view, since standard evolutionary biology is clearly not a mechanistic science: it does not explain evolutionary changes by decomposing a system into its parts and activities and recomposing it back. Rather, it takes the form of a statistical explanation. However, claims about the causal impact of development in evolution must be taken seriously from a causalist position too. Thus, one shall consider that development has a role in these population-level causes supposedly represented in statistical models. Advocates of the mechanistic view of evolution try to understand population-level evolutionary causes better by fitting them into the mechanistic framework. Thus, the evo-devo mechanistic question is: is there a way in which development is relevant *qua* evolutionary mechanism?

In population-level mechanisms of evolution, specific causal chains are *instantiations* of the mechanism as a *type*. For example, specific ecological processes that explain particular adaptations are instantiations of natural selection as a type of mechanism (Skipper & Millstein, 2005). Similarly, we need to think of developmental biases and innovation not as the result of specific causal chains in lineages—as those explained in terms of lineage explanations (Calcott, 2009)—but as a *type* of evolutionary phenomena instantiated in different causal chains. Recall that the evo-devo agendas on homology provide general views of innovation, either by pointing at the co-option of a mechanistic component into a different character

²However, it is not clear that reproduction counts as an organismal process, as it necessarily involves at least two organisms. Although this might hinder a purely organism-centered view of evo-devo (see Villegas & Triviño, 2023), it does not affect the arguments of this chapter.

mechanism (DiFrisco et al., 2023) or by stressing the stages of generation, integration and autonomization of a new trait (Müller, 2003). These mechanistic structures may be general enough to qualify as a *type* of evolutionary mechanism in the population-level sense. For example, the generation of a new phenotypic component through epigenetic processes can take place in any given population. However, these proposals still refer to particular kinds of developmental bias, and particularly of innovation, but not all kinds of phenotypic change seem to be explainable in their terms.

Both classical and recent work on evo-devo mechanisms can cast light into the mechanistic picture of evolution through the broad notion of “developmental repatterning” (Arthur, 2011). Developmental repatterning is a term used to refer to generalizable changes in development that produce evolutionary changes, such as heterochrony, heterotopy, heterometry and heteronomy (or heterocyberny, see Moczek, 2019). These processes make reference to changes in timing, location, amount or nature, respectively, of a component in a developmental mechanism that produces a phenotypic change. Classical evo-devo work is interested in these phenomena as a developmental kind of *evolutionary mechanism* (Hall, 2003), although the literature on evo-devo mechanisms has not dealt in detail with this broader idea of mechanism. It is important to stress that these developmental phenomena are identified in the literature as mechanisms of evolution, as exemplified by heterochrony as “a mechanism for evolutionary diversification of flower form” (Endress, 2006, p. 5). For example, changes in the timing of sepal production in the floral meristem explain variation in the size and number of sepals in the flowers of *Dipsacoidae* species (Naghiloo & Claßen-Bockhoff, 2017).

The kinds of phenotypic changes produced by developmental repatterning can be gradual, such as the size and number variation in sepals just mentioned, but they can be evolutionary novelties too. For example, heterotopy of B-type and C-type gene expression seems to have been involved in the origin of the flower plan. These genes are associated with male and female organs, respectively, in the ancestral lineage of angiosperms, and their conjunct expression in the same axes may have resulted in the origin of the bisexual plan of flowers (Wagner, 2014). Both the small heterochronic changes producing sepal size variation and the greater heterotopic changes involved in the origin of flowers are specific instances of developmental repatterning.

In turn, developmental repatterning refers to an abstract mechanistic explanation of how phenotypic variants and novelties arise in populations. Therefore, it can be used in the same sense that selection, drift and mutations can be thought of as mechanisms: general mechanistic structures that are alluded to for explaining populational phenomena in evolutionary explanations, and that can be described mechanistically only in the context of other explanatory agendas with a focus on organism-level phenomena. In these organism-level mechanistic explanations, scientists are not explaining *the* mechanism of, say, heterochrony, just like ecological mechanistic explanations don't explain *the* mechanism of natural selection. Rather, they explain the mechanism for a particular heterochronic change in a lineage (such as the heterochronic change in sepal development), or a particular episode of selection in a population.

For developmental repatterning to count as a mechanism, it must be accountable in terms of entities and activities organized in a certain way. Here I provide a very minimal characterization. At the very least, two individuals forming a lineage must be involved. Importantly, their genotypes, phenotypes, and developmental mechanisms are constitutive parts of developmental repatterning, and so are their own mechanistic components at lower levels. The relevant activities for developmental repatterning are reproduction, development, mutations and recombinations. These entities and activities are organized in such a way that reproduction of organisms generates new phenotypic variation biased by the properties of the underlying developmental mechanisms. Biased phenotypic changes functionally individuate developmental repatterning as a mechanism, similarly to how enhanced survival and reproduction of a type individuate selection. Again, the specific phenomenon produced can be very variable, ranging from minor changes to phenotypic novelties³, just like selection can lead to the stabilization of a trait or to the emergence of complex adaptations. The difference will lay in how reproduction combines the mechanistic elements present in the lineage given the inputs it receives from mutations, recombinations, or environmental elements.

How does developmental repatterning relate to other population-level mechanisms of evolution? In mechanistic views of selection, phenotypic variation has tended to be considered either as a temporal stage (Skipper & Millstein, 2005) or as a component entity in a lax sense of natural selection (Illari & Williamson, 2010). This reflects the fact that variation is a precondition for evolution by natural selection to occur. As such, it tends to be assumed whenever mechanistic accounts of selection are discussed: it is *variation* in the way individual organisms deal with the environment that allow for differential survival in populations. However, this treatment ignores another fact about variation, namely that it is *produced* in iterations of reproduction independently of natural selection. The multilevel mechanistic proposal of selection (Illari & Williamson, 2010) is an exception, interpreting phenomena such as recombination or epistasis (i.e., gene expression that depends on the presence of other genes) as part of what constitutes selection. Considering these development-related phenomena as part of natural selection is nevertheless problematic. They are not part of selection but of the phenotypic *response* of the population to an episode of selection. That is, they are part of the way the reproductive and developmental properties of the population provide new variation once an episode of natural selection has occurred. Here is where developmental repatterning enters the scene.

Responses to selection depend on the genotype-phenotype structure, namely the way genotypic variation maps into phenotypic one. Although there is debate over whether the genotype-phenotype structure has evolved such that it promotes or facilitates adaptation, developmental organization is involved in producing all kinds of variation, adaptive or not. A population that changes through drift or mutations

³ Including novelties driven by co-option of lower-level mechanistic elements (DiFrisco et al., 2023) or by epigenetic processes (Müller, 2003).

will also generate (sometimes new) variation mediated by the genotype-phenotype structure. Thus, the developmental repatterning responsible for phenotypic responses cannot be solely a constitutive of any of the other population-level mechanisms of evolution such as selection or drift. Instead, it is one mechanistic component of evolutionary change acting in conjunction with other population-level mechanisms. Given the ongoing nature of cycles of differential reproduction and generation of variation that evolution consists of, developmental repatterning can be seen as a previous mechanistic step for natural selection or drift, or as the step following them and mutations. In any case, it is a distinct mechanism that shall not be conflated with other population-level evolutionary mechanisms.

4.5 Concluding Remarks

We have seen that a great deal of the evo-devo agenda is mechanistic, especially when it comes to individuating developmental systems phylogenetically through the study of homology and evolutionary novelties. Here, it seems that there is an inclination in evo-devo to regard causal and constitutive relations of developmental elements and activities as explanatory of phenotypic unity and diversity. In this sense, there is a clear causal aspect of developmental mechanisms in the production of evolutionarily relevant variation in specific lineages. However, bringing mechanistic components to general evolutionary explanations and predictions is a different story. Mechanistic knowledge of developmental systems needs to be used in combination with other means of explanation, mostly topological and dispositional, sometimes to the extent that there is little mechanistic content in some evo-devo explanations of phenotypic change (e.g., through statistical uses of the genotype-phenotype map). This is not a limitation of the evo-devo approach, but a very interesting point of connection with more classical approaches to evolution. While it is obviously concerned with the mechanistic aspects of development that are relevant to evolution, the field is growing significantly in its incorporation of developmental biases into population-level studies of evolution. In doing so, it makes a more general statement about the causal impact of development that is not always explained in terms of specific developmental mechanisms.

This situation may seem to imply that it is not the mechanistic aspect of evo-devo what justifies the causal role of development that it forcefully vindicates, and therefore that the two lemmas of evo-devo are not directly related (cf. Wagner et al., 2000; Hall, 2003; Müller, 2007). However, in this chapter I have provided an alternative view through the characterization of developmental repatterning as a population-level evolutionary mechanism. Previous views about evolutionary mechanisms failed to articulate the relation that variation holds with other evolutionary factors understood as mechanisms. Developmental repatterning provides a mechanistic view of the generation of variation that acts in combination with other mechanisms of evolution. The generality of developmental repatterning as a mechanistic structure means that it is not restricted to the impact of specific developmental

mechanisms in a particular lineage—as it is often thought to be the main contribution of evo-devo. Rather, it refers to the organizational properties of all lineages that, through reproduction, development, mutations and recombinations, channel phenotypic changes through the properties of developmental mechanisms. I believe that incorporating developmental repatterning as the mechanism for phenotypic change and evolutionary novelty into the broader picture of evolutionary mechanisms helps situate better the agenda of evo-devo and its vindications on the causal role of development into our philosophical discussions of evolution.

Acknowledgements I thank the book editors for their kind invitation to contribute to this volume. Davide Vecchi, Laura Nuño de la Rosa and an anonymous reviewer provided very valuable comments on previous versions of this chapter. I also thank audiences at the University of Lisbon (CFCUL and ReFiCi joint seminar, and the Workshop ‘Mechanism and its Enemies’) and at the Australia/New Zealand Philosophy of Biology Workshop (ANU, Canberra) for their feedback. I am grateful to the Konrad Lorenz Institute for Evolution and Cognition Research (KLI, Austria) for support at the early stages of this research. This work was funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., in the R&D Center for Philosophy of Sciences of the University of Lisbon (CFCUL), strategic project with references FCT I.P. UIDB/00678/2020 and UIDP/00678/2020, and through the contract with reference 2021.03186.CEECIND/CP1654/CT0008.

References

- Alvarez-Buylla, E. R., Benítez, M., Corvera-Poiré, A., Cador, Á. C., de Folter, S., de Buen, A. G., Garay-Arroyo, A., García-Ponce, B., Jaimes-Miranda, F., Pérez-Ruiz, R. V., Piñeyro-Nelson, A., & Sánchez-Corrales, Y. E. (2010). Flower development. *The Arabidopsis Book*, 8, e0127.
- Amundson, R. (2015). Preface. In A. Love (Ed.), *Conceptual change in biology* (pp. v–x). Springer.
- Arthur, W. (2011). *Evolution: A developmental approach*. Wiley.
- Austin, C. J. (2016). The ontology of organisms: Mechanistic modules or patterned processes? *Biology & Philosophy*, 31(5), 639–662.
- Austin, C. J. (2017). Evo-devo: A science of dispositions. *European Journal for Philosophy of Science*, 7(2), 373–389.
- Austin, C. J., & Nuño de la Rosa, L. (2021). Dispositional properties in evo-devo. In L. Nuño de la Rosa & G. Müller (Eds.), *Evolutionary developmental biology: A reference guide* (pp. 469–481). Springer.
- Baedke, J. (2021). Mechanisms in evo-devo. In L. Nuño de la Rosa & G. Müller (Eds.), *Evolutionary developmental biology: A reference guide* (pp. 383–395). Springer.
- Baedke, J., & Mc Manus, S. F. (2018). From seconds to eons: Time scales, hierarchies, and processes in evo-devo. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 72, 38–48.
- Baetu, T. M. (2012). Mechanistic constraints on evolutionary outcomes. *Philosophy of Science*, 79(2), 276–294.
- Barros, D. B. (2008). Natural selection as a mechanism. *Philosophy of Science*, 75(3), 306–322.
- Bateson, P., & Laland, K. N. (2013). Tinbergen’s four questions: An appreciation and an update. *Trends in Ecology & Evolution*, 28(12), 712–718.
- Baum, D. A., & Hileman, L. C. (2006). A developmental genetic model for the origin of the flower. *Annual Plant Reviews: Flowering and its Manipulation*, 20, 1–27.
- Beatty, J. (1984). Chance and natural selection. *Philosophy of Science*, 51(2), 183–211.

- Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science Part A*, 41(3), 321–333.
- Bechtel, W., & Abrahamsen, A. (2013). Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations. *From Ontos Verlag: Publications of the Austrian Ludwig Wittgenstein Society-New Series (Volumes 1–18)*, 12.
- Bell, G. (2008). *Selection: The mechanism of evolution*. Oxford University Press on Demand.
- Bich, L., & Skillings, D. (2023). There are no intermediate stages: An organizational view on development. In M. Mossio (Ed.), *Organization in biology* (pp. 241–262). Springer.
- Brigandt, I. (2015). Evolutionary developmental biology and the limits of philosophical accounts of mechanistic explanation. In P. Braillard & C. Malaterre (Eds.), *Explanation in biology* (pp. 135–173). Springer.
- Calcott, B. (2009). Lineage explanations: Explaining how biological mechanisms change. *The British Journal for the Philosophy of Science*, 60(1), 51–78.
- Chartier, M., Jabbour, F., Gerber, S., Mitteroecker, P., Sauquet, H., von Balthazar, M., Staedler, Y., Crane, P. R., & Schoenenberger, J. (2014). The floral morphospace – A modern comparative approach to study angiosperm evolution. *New Phytologist*, 204(4), 841–853.
- Conley, B. A. (2020). Mayr and Tinbergen: Disentangling and integrating. *Biology & Philosophy*, 35(1), 1–23.
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.
- Craver, C. F., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.
- DesAutels, L. (2016). Natural selection and mechanistic regularity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 57, 13–23.
- DesAutels, L. (2018). Mechanisms in evolutionary biology. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 296–307). Routledge.
- DiFrisco, J., Love, A. C., & Wagner, G. P. (2020). Character identity mechanisms: A conceptual model for comparative-mechanistic biology. *Biology & Philosophy*, 35(4), 1–32.
- DiFrisco, J., Wagner, G. P., & Love, A. C. (2023). Reframing research on evolutionary novelty and co-option: Character identity mechanisms versus deep homology. *Seminars in Cell & Developmental Biology*, 145, 3–12.
- Endress, P. K. (2006). Angiosperm floral evolution: Morphological developmental framework. *Advances in Botanical Research*, 44, 1–61.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342–S353.
- Hall, B. K. (2003). Evo-devo: Evolutionary developmental mechanisms. *International Journal of Developmental Biology*, 47(7–8), 491–495.
- Huneman, P. (2018). Diversifying the picture of explanations in biological sciences: Ways of combining topology with mechanisms. *Synthese*, 195(1), 115–146.
- Hüttemann, A., & Kaiser, M. I. (2018). Potentiality in biology. In K. Engelhard & M. Quante (Eds.), *Handbook of potentiality* (pp. 401–428). Springer.
- Illari, P. M., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 41(3), 279–291.
- Jaeger, J., & Sharpe, J. (2014). On the concept of mechanism in development. In A. Minelli & T. Pradeu (Eds.), *Towards a theory of development* (pp. 56–78). Oxford University Press.
- Kaiser, M. I. (2021). Explanation in evo-devo. In L. Nuño de la Rosa & G. Müller (Eds.), *Evolutionary developmental biology: A reference guide* (pp. 357–370). Springer.
- Laland, K. N., Sterelny, K., Odling-Smee, J., Hoppitt, W., & Uller, T. (2011). Cause and effect in biology revisited: Is Mayr's proximate-ultimate dichotomy still useful? *Science*, 334(6062), 1512–1516.

- Leuridan, B., & Lodewyckx, T. (2021). Diachronic causal constitutive relations. *Synthese*, 198(9), 9035–9065.
- Love, A. C. (2018). Developmental mechanisms. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 332–347). Routledge.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Mayr, E. (1961). Cause and effect in biology. *Science*, 134, 1501–1506.
- Mc Manus, F. (2012). Development and mechanistic explanation. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(2), 532–541.
- Millstein, R. L. (2003). Interpretations of probability in evolutionary theory. *Philosophy of Science*, 70(5), 1317–1328.
- Minelli, A., & Pradeu, T. (Eds.). (2014). *Towards a theory of development*. Oxford University Press.
- Moczek, A. P. (2019). The shape of things to come: Evo devo perspectives on causes and consequences in evolution. In T. Uller & K. N. Laland (Eds.), *Evolutionary causation: Biological and philosophical reflections* (pp. 23–63). MIT Press.
- Mondragón-Palomino, M., Hiese, L., Härter, A., et al. (2009). Positive selection and ancient duplications in the evolution of class B floral homeotic genes of orchids and grasses. *BMC Ecology and Evolution*, 9, 81.
- Müller, G. B. (2003). Homology: The evolution of morphological organization. In G. Müller & S. Newman (Eds.), *Origination of organismal form: Beyond the gene in developmental and evolutionary biology* (pp. 51–70). The MIT Press.
- Müller, G. B. (2007). Six memos for evo-devo. In M. Laubichler & J. Maienschein (Eds.), *From embryology to evo-devo: A history of developmental evolution* (pp. 499–524). The MIT Press.
- Müller, G. B., & Newman, S. A. (2005). The innovation triad: An EvoDevo agenda. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 304(6), 487–503.
- Naghiloo, S., & Claßen-Bockhoff, R. (2017). Developmental changes in time and space promote evolutionary diversification of flowers: A case study in Dipsacoideae. *Frontiers in Plant Science*, 8, 1665.
- Newman, S. A. (2003). From physics to development: The evolution of morphogenetic mechanisms. In G. Müller & S. Newman (Eds.), *Origination of organismal form: Beyond the gene in developmental and evolutionary biology* (pp. 221–240). The MIT Press.
- Newman, S. A. (2014). Physico-genetics of morphogenesis: The hybrid nature of developmental mechanisms. In A. Minelli & T. Pradeu (Eds.), *Towards a theory of development* (pp. 95–113). Oxford University Press.
- Nuño de la Rosa, L. (2018). Capturing processes: The interplay of modelling strategies and conceptual understanding in developmental biology. In D. Nicholson & J. Dupré (Eds.), *Everything flows* (pp. 264–283). Oxford University Press.
- Nuño de la Rosa, L., & Etxeberria, A. (2012). Pattern and process in evo-devo: Descriptions and explanations. In *EPSA philosophy of science: Amsterdam 2009* (pp. 263–274). Springer.
- Nuño de la Rosa, L., & Villegas, C. (2022). Chances and propensities in evo-devo. *The British Journal for the Philosophy of Science*, 73(2), 509–533.
- Pavličev, M., Bourg, S., & Le Rouzic, A. (2023). The genotype-phenotype map structure and its role for evolvability. In T. Hansen, D. Houle, M. Pavličev, & C. Pélabon (Eds.), *Evolvability: A unifying concept in evolutionary biology?* (pp. 147–170). The MIT Press.
- Pence, C. H. (2021). *The causal structure of natural selection*. Cambridge University Press.
- Pérez-González, S., & Luque, V. J. (2019). Evolutionary causes as mechanisms: A critical analysis. *History and Philosophy of the Life Sciences*, 41(2), 1–23.
- Peterson, T., & Müller, G. B. (2016). Phenotypic novelty in EvoDevo: The distinction between continuous and discontinuous variation and its importance in evolutionary theory. *Evolutionary Biology*, 43(3), 314–335.
- Pigliucci, M., & Kaplan, J. (2010). *Making sense of evolution: The conceptual foundations of evolutionary biology*. University of Chicago Press.

- Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese*, 192(11), 3731–3755.
- Skipper, R. A., & Millstein, R. L. (2005). Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 327–347.
- Theißen, G., & Rümpler, F. (2021). Evolution of floral organ identity. In L. Nuño de la Rosa & G. Müller (Eds.), *Evolutionary developmental biology: A reference guide* (pp. 697–714). Springer.
- Theißen, G., Melzer, R., & Rümpler, F. (2016). MADS-domain transcription factors and the floral quartet model of flower development: Linking plant development and evolution. *Development*, 143(18), 3259–3271.
- Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, 20, 410–433.
- Villegas, C., & Triviño, V. (2023). Typology and organismal dispositions in evo-devo: A meta-physical approach. *ArtefaCToS. Journal of Science and Technology Studies*, 12(1), 79–103.
- Wagner, G. P. (2014). *Homology, genes, and evolutionary innovation*. Princeton University Press.
- Wagner, G. P., Chiu, C. H., & Laubichler, M. (2000). Developmental evolution as a mechanistic science: The inference from developmental mechanisms to evolutionary processes. *American Zoologist*, 40(5), 819–831.
- Walsh, D. M., Ariew, A., & Matthen, M. (2017). Four pillars of statisticalism. *Philosophy, Theory, and Practice in Biology*, 9(1).
- Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, 78(2), 277–297.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 5

Organisms Need Mechanisms; Mechanisms Need Organisms



William Bechtel and Leonardo Bich

Abstract According to new mechanists, mechanisms explain how specific biological phenomena are produced. New mechanists have had little to say about how mechanisms relate to the organism in which they reside. A key feature of organisms, emphasized by the autonomy tradition, is that organisms maintain themselves. To do this, they rely on mechanisms. But mechanisms must be controlled so that they produce the phenomena for which they are responsible when and in the manner needed by the organism. To account for how they are controlled, we characterize mechanisms as sets of constraints on the flow of free energy. Some constraints are flexible and can be acted on by other mechanisms, control mechanisms, that utilize information procured from the organism and its environment to alter the flexible constraints in other mechanisms so that they produce phenomena appropriate to the circumstances. We further show that control mechanisms in living organisms are organized heterarchically—control is carried out primarily by local controllers that integrate information they acquire as well as that which they procure from other control mechanisms. The result is not a hierarchy of control but an integrated network of control mechanisms that has been crafted over the course of evolution.

Keywords Autonomy · Control mechanisms · Constraints · Free energy, heterarchical organization · Mechanistic explanation

W. Bechtel (✉)

Department of Philosophy, University of California, San Diego, La Jolla, CA, USA
e-mail: wbechtel@ucsd.edu

L. Bich

IAS-Research Centre for Life, Mind and Society, Department of Philosophy,
University of the Basque Country (UPV/EHU), Donostia-San Sebastian, Spain
e-mail: leonardo.bich@ehu.es

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of
the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_5

5.1 Introduction

Among the entities in the universe, organisms are highly unusual. They are complexly organized systems made of soft materials that tend to degrade, yet they maintain themselves far from equilibrium. This requires regular work—an organism must extract free energy and materials from the environment and utilize them to construct, repair, and maintain itself. When organisms stop performing this work, they die and decay (generally assisted by other organisms that use the matter and energy accessed from the dead organism for their own self maintenance). Although all individual organisms eventually die, all organisms now alive are parts of continuous lineages of organisms which, over a span of more than three billion years since the origin of life, maintained themselves and produced successors.

To perform the work needed to maintain themselves, organisms rely on mechanisms—sets of components organized to carry out different activities in a coordinated fashion. As envisaged by some new mechanists, mechanisms are active—according to Machamer et al. (2000), a mechanism produces a specific phenomenon whenever its start-up conditions are realized.¹ What phenomenon? The phenomenon the mechanism is equipped by its constitution to produce. These phenomena (e.g., protein synthesis, generating action potentials, cell division) are far less complex than life itself. In advancing their account, mechanists have had little to say about whole organisms and how they act to maintain themselves. They seem to treat organisms as simply collections of mechanisms. If mechanisms are construed, as they are by the new mechanists, as each responsible for one phenomenon, then the organism must consist of just the right set of mechanisms to generate each phenomenon when it is needed so that the organism is maintained. Given the constantly changing conditions that organisms confront, it is extremely unlikely that even a powerful process such as evolution by natural selection could have equipped organisms with just the right set of single-phenomenon mechanisms to jointly execute the actions organisms need to survive. This seems even less likely when one recognizes that organisms are also agents which change their environments and thereby alter what activities they must perform to maintain themselves. A minimal step to overcoming this challenge is to reconceptualize mechanisms so that they are capable of producing different phenomena as required by the circumstances an organism finds itself in. In Sect. 5.2, we offer a revisionist account that characterizes mechanisms in terms of constraints on flows of free energy and show how it provides for a more dynamical account in which mechanisms can be controlled so as to perform different activities as needed.

A different tradition of theorists, constituting the organization or autonomy school, has focused directly on the ability of organisms to maintain themselves. Theorists in this tradition address such topics as the ability of organisms to construct themselves (Maturana & Varela, 1980), repair themselves (Rosen, 1991), and

¹Bechtel and Abrahamsen (2009); (see also Abrahamsen & Bechtel, 2011) also view mechanisms as capable of endogenous activity, but for them that is a consequence of cyclic organization.

manage thermodynamic processes (Moreno & Mossio, 2015). This tradition treats organisms as the active entities—they act to maintain themselves. It is, however, challenging to explain how organisms themselves have such capacities: what is the organism to which such actions are attributed? The organism is the whole organized system comprising its various components. It is not something additional to its constituents that has its own powers (Ryle, 1949). To explain how the organism carries out any given activity needed to maintain itself, scientists appeal to its component mechanisms and what they do. Each activity an organism performs results from the operation of specific mechanisms in it. In what sense is the organism, not the component mechanisms, responsible for carrying out the appropriate activities?² Here the autonomy school³ offers an important insight—each mechanism that carries out an activity needed for the organism to maintain itself is the product of closed loops of processes within the organism. Different theorists characterize how these processes are closed in different terms. For Maturana and Varela, it is closure of construction (autopoiesis), for Rosen, closure of efficient causation, and for Moreno and Mossio, closure of constraints. We discuss the different conceptions of closure in Sect. 5.3, showing that the notion of closure of constraints offers the greatest promise for understanding organisms as maintaining themselves.

Having characterized both mechanisms and autonomy in terms of constraints, we explore how a focus on constraints can serve to integrate these perspectives in Sect. 5.4. The key to doing this is developing an account of how mechanisms can act on the constraints of other mechanisms. We treat mechanisms that act on the constraints of other mechanisms as belonging to a distinct type of mechanism, *control mechanisms*. Most of the mechanisms characterized by the new mechanists are what we term *production mechanisms*—mechanisms that constrain free energy to carry out a productive activity—constructing something, moving it about, or taking it apart. On our conception of mechanisms, mechanisms perform specific productive activities due to how constraints realized in them direct flows of free energy. Control mechanisms, as we understand them, also operate as a result of constraining flows of free energy, but do so to modify constraints in other mechanisms, thereby determining how those mechanisms operate.⁴ Thus, control mechanisms direct the activities of production mechanisms. If closure of constraints is to explain how organisms are capable of acting to maintain themselves, then we must characterize how the constraints in control mechanism are part of the closed system. This is tricky since, for control mechanisms to do their job, the constraints realized in them at a given time need to be responsive to conditions external to themselves. We will

²Each composite of mechanisms can, of course, produce different activities. This is due to how the components are organized.

³Another term applied to the autonomy school is the organization school. The key thing to emphasize is that the activities of an organized system are due to the components acting together within the overall organization.

⁴A notable feature of control mechanisms is that generally they require much less energy than production mechanisms. It takes much less free energy to move a switch than it does to operate a motor.

develop an account of how control mechanisms can be open to information they procure from the environment and yet part of a closed network in the sense required for their activities to be viewed as activities of the organism.

In thinking about control, especially in the design of social and political institutions, we often think hierarchically — local controllers report to a smaller set of controllers at the next level and at the top is the chief executive that controls the whole institution. This perspective is sometimes adopted in thinking about biological organisms—theorists might conceptualize the nervous system as controlling an organism’s body and as itself organized hierarchically such that higher centers in the brain control others. This, however, misrepresents how control mechanisms in biological organisms are organized—much control remains local and multiple local controllers coordinate their activities to accommodate the diverse needs of the organism without any one controller being in charge. We develop this understanding of heterarchical control in Sect. 5.5 before concluding in Sect. 5.6.

5.2 Constraints: A Revisionist Account of Mechanisms

The new mechanists articulated their concept of mechanism as they sought to understand the practices of biologists who frequently appeal to mechanisms in their explanations of biological phenomena (Bechtel & Richardson, 1993/2010). Taking their lead from the fact that biologists develop their mechanistic accounts by decomposing systems taken to be responsible for a phenomenon into their constituents, the new mechanists characterized mechanisms in terms of constituent entities or parts,⁵ the activities or operations performed by these entities, and how they are organized to produce the phenomenon (Machamer et al., 2000; Bechtel & Abrahamsen, 2005; Glennan, 2017). In developing their account, Machamer et al. chose the term *activity* to emphasize that mechanisms do things and insisted on a dualism of entities and activities. On their construal, the activity of a whole mechanism results from the activities of their component entities. This seems to introduce a regress in which to explain any activity one must decompose a given mechanism into its component activities. Machamer et al. seek to stop the regress by noting that in practice the explanations researchers advance bottom-out with components whose activities are simply accepted and not further explained. Whether one terminates the decomposition at a given point or continues further down, the notion of activity remains a primitive that is not itself explained (in particular, on their view, it is not explained by the entities that constitute the mechanism).

⁵Philosophers of science who have advanced a process metaphysics (see various contributions to Nicholson & Dupré, 2018) have criticized the mechanists appeal to entities or parts, construing the mechanists as treating these as unchanging *things*. It is important to recognize that new mechanists do not view mechanisms or their components entities or parts as unchanging. The entities constituting mechanisms change as mechanisms operate and as they are constructed, repaired, and eventually deconstructed.

A different approach is to appeal to how physicists explain changes in the world. They commonly appeal to Gibbs free energy. Although matter and energy are viewed as interconvertible, the differentiation of matter and energy itself represents a dualism. It is a dualism, however, that is grounded in basic laws thought to govern the universe. According to thermodynamics, maximal free energy was available at the origin of the universe when matter was unequally distributed and dissipates as matter becomes more homogeneously distributed. The second law of thermodynamics asserts that in any closed system, free energy continually dissipates as the distribution of matter goes to equilibrium. Available energy due to disequilibrium within a system can be used to produce mechanical work. It can only do so, however, when it is constrained—left unconstrained, free energy is lost as the system goes to thermal equilibrium.

Thus, the key to the ability of a system such as a mechanism to perform work is that free energy does not simply dissipate but does so in a constrained manner (Kauffman, 2000). The notion of constraint was introduced into classical mechanics to account for macroscale objects. On its own, each elementary particle can move in any of six dimensions (three spatial, three rotational). But when these particles are bound to each other (e.g., through chemical bonds), they are constrained to move with the composite object. When a force is applied to the composite object, its components move with it due to the constraints. The notion of constraint can be extended to thermodynamics: how free energy dissipates is constrained by the current structure of the system. For example, when free energy is released by combustion in the cylinder of a gasoline engine, it is constrained to move against the piston, thereby performing work (Hooker, 2013).

A focus on constraints as physical structures that limit the flow of free energy is crucial in understanding how biological organisms direct free energy into the production of work. We cannot provide a thorough discussion of constraints here, but note two important features of constraints. First, in conceptualizing a structure as a constraint, one needs to specify the time-scale during which it serves as a constraint. As physical structures, constraints can and are changed by flows of free energy. Distinguishing a constraint from the process of energy flowing through the constraints to produce work, Mossio et al. (2013) contend that at the time-scale characteristic of the process, the constraint is locally unaffected by the process—the constraint is not part of the process and is stable during it. Moreover, at that time-scale, the constraint exerts a distinctive causal power on the process, limiting the range of possible outcomes (degrees of freedom) of the process. Second, although the term *constraint* emphasizes that constraints impose limits, Kauffman (2000) and Hooker (2013) among others, have developed how constraints are also enabling as they create new possibilities. By canalizing the flow of free energy, constraints enable outcomes that otherwise would be extremely improbable or practically impossible. When water flows downhill, free energy is dissipated. But if it is limited to flowing through a pipe, the water in a reservoir can reach a distant tank that it would not otherwise reach. As a result of the pipe, water molecules that might have flowed in different direction are limited to following in the same direction. If further constraints are added, this directed flow of water can be used to carry out other

activities, such as moving the wheel of a water wheel, resulting in the milling of grain that would not otherwise be turned into flour. Hooker also provides an illustrative biological example: a skeleton restricts the movements an organism can make, but also enables it to move in ways it couldn't otherwise.

Winning and Bechtel (2018) adapted this perspective on free energy, constraints, and work to characterizing mechanisms. They viewed the components of mechanisms as imposing constraints restricting the flow of free energy. On this view, biological mechanisms are active not because they are composed of activities, but because they constrain free energy so as to perform work—to generate the phenomenon for which the mechanism is taken to be responsible. On the conception of mechanism proposed by Winning and Bechtel, mechanisms should not be understood simply as organized sets of entities and activities, but as organized sets of constraints (entities, parts) that direct the flow of available free energy so as to carry out work (generate the phenomenon). The notion of activity (or operation) still has a place in this account—as researchers decompose the mechanism in their attempt to understand how it generates the phenomenon, they will focus on the activities of individual components of the mechanism. These activities, however, will not be treated as primitives, but as the product of the constraint of free energy by particular components of the mechanism.

Although philosophical accounts of mechanism prior to Winning and Bechtel did not attend to the role of free energy in mechanisms, it has clearly been central to biological thinking since the pioneering work of Lavoisier and Laplace (1780), who characterized the metabolic activities of animals in terms of combustion. One of the most prominent physiological chemists of the nineteenth century, Liebig (1840), sharply distinguished between plants as synthesizing energy rich molecules such as sugars, and animals as acquiring energy by catabolizing them. Although this simple assignment of synthesis to plants and catabolism to animals was soon recognized as too simplistic as animals also carry out synthesis, physiologists focused on heat as the energy currency of animals (Mendelsohn, 1964). This changed with the discovery that adenosine triphosphate (ATP) provided the free energy for animal activities such as muscle contraction (Fiske & Subbarow, 1929; Lohmann, 1929). Due to the unusual amount of free energy liberated by the hydrolysis of ATP to adenosine diphosphate (ADP), the bond to the third phosphate group came to be regarded as a “high-energy” bond and the primary energy currency in animals. Initially, physiologists could do little more than correlate ATP synthesis with the catabolism of sugars, fatty acids, and other molecules and the hydrolysis of ATP with activities such as muscle contraction. For example, after (Huxley, 1969) advanced the swinging-crossbridge model of how myosin exerted force on actin in the course of muscle contraction, Lynn and Taylor (1971) associated each step with a step in the process of ATP hydrolysis. By the 1990s, though, researchers began to explicate this process in terms of the chemical bonds formed between a substrate and ATP that enabled the energy liberated in hydrolysis to be constrained within myosin so as to move another part of the molecule, referred to as the lever arm, whose movement exerted force on actin (Fisher et al., 1995; Holmes & Geeves, 2000; for theoretical analysis, see Bechtel & Bollhagen, 2021). Similar analysis of the molecules involved in ATP

synthesis showed how free energy captured in a proton gradient in the mitochondria generated force within the F_0F_1 ATPase that brings ADP and Pi into juxtaposition so that they form a bond.

The ability to analyze the flow of free energy in terms of forces exerted in molecular structures is still only possible in limited cases.⁶ In many cases, physiologists can only appeal generally to the role of ATP in supplying the source of free energy. This is especially true at higher levels of organization in which researchers characterize the activity of muscle in phenomena such as the pumping of the heart or how foodstuffs are broken down and transferred through the organs of the digestive tract. Although they cannot show in detail how the energy released by hydrolysis of ATP is constrained so as to create force that results in the physical work, they nonetheless frequently identify where hydrolysis occurs that provides the needed free energy for a given mechanism to operate. What the revisionist account of mechanism makes clear is that what the biologists are envisaging is constraints restricting the flow of free energy through mechanisms.

An important benefit provided by the revisionist account in contrast to standard new mechanist accounts (e.g., Machamer et al., 2000; Bechtel & Abrahamsen, 2009; Glennan, 2017) is that it makes clear how mechanisms are dynamic, capable of varying their operation and even carrying out multiple activities. Most new mechanists have embraced what has been referred to as Glennan's law which identifies one mechanism with one phenomenon.⁷ Bollhagen and Bechtel (2022) have shown that in practice, once researchers have used the characterization of a phenomenon to pick out a mechanism, they anchor their further investigations on the mechanism itself. This sometimes leads to discovering that the same mechanism is responsible for different phenomena. For example, it is not uncommon that, after discovering the mechanism responsible for a phenomenon, researchers determine that it often autoinhibits—prevents itself from operating except when conditions require its operation. This is made possible by the fact that not all the constraints constituting a mechanism are fixed. Some can be acted on and changed.

Even the production of the initially characterized phenomenon typically requires that constraints within the mechanism be changed as energy is directed through the mechanism. For example, in a human-made machine such as a car engine, the piston moves as a result of the free energy released through the combustion of gasoline. Pistons are connected through the camshaft so that, as one piston moves, it applies force to others. Among other things, this compresses the gasoline in another cylinder. Once compressed, a spark initiates its combustion, which acts on the first piston, returning it to its original position to begin another cycle of activity. A similar

⁶This is changing rapidly. See, for example, Swan et al. (2021) for an account of how ATP hydrolysis generates movements within KaiC that provides the free energy for the cycle of events that constitute a circadian cycle in cyanobacteria.

⁷Glennan (1996) argued that "One cannot even identify a mechanism without saying what it is that the mechanism does." An exception to the widespread endorsement of this contention is Bechtel and Abrahamsen's (2005, p. 423) acknowledgment that a mechanism may be "responsible for one or more phenomena."

cycle of changing flexible constraints figures in the action of myosin—the hydrolysis of ATP results in changing the constraints within myosin, altering its ability to bind actin and to exert force on it, culminating in it expelling ADP and binding a new molecule of ATP. In these processes, constraints result in movement that changes the constraints, altering subsequent movement.

In addition to being changed in the normal working of a mechanism, constraints can be changed by other mechanisms working on it. By changing the constraints in a mechanism, these other mechanisms can change how free energy flows through the first mechanism and thus what work it performs. To illustrate this, we return to the case of actin and myosin. By default, the sites at which myosin can bind actin are blocked by tropomyosin binding to them. When calcium ions (Ca^{2+}) are released into the cytoplasm, they bind tropomyosin and remove it from the myosin binding site. Normally whatever Ca^{2+} is in the cytoplasm is taken up in the sarcoplasmic reticulum, but when signaling proteins bind receptors on the sarcoplasmic reticulum, they change constraints in those receptors, allowing Ca^{2+} to flow into the cytoplasm and remove tropomyosin, allowing myosin to bind and exert force on actin.

In Sect. 5.4 we will characterize mechanisms that change constraints within other mechanisms as control mechanisms. But first we turn to the autonomy tradition, which has also foregrounded the notion of constraint and made it central to the account of closure that renders organisms autonomous.

5.3 Autonomy and the Closure of Constraints

Kant (1790/1987) famously advanced the idea that organisms are self-determining—are autonomous. This meant that some of the causes of the existence of an organism are not external and independent from it, but depend on the very organism that they help to generate. Another way to state the Kantian idea is that the system and its components are mutually dependent, as the components exist for the whole they generate and the whole exists for the components it produces and maintains. The challenge is to work out just what this entails. Piaget (1967), Rosen (1972), and Maturana and Varela (1980), among others, emphasized that organisms are systems organized in such a way that they are capable of constructing, repairing, and maintaining their parts, and consequently themselves, through the continuous exchange of matter and energy with the environment—they are autopoietic. Insofar as the functional components responsible for these activities are made by the organisms themselves (or by their predecessors), what the components do can be viewed derivatively as activities of the organism. Maturana (1980, p. 48) comments “The living organization is a circular organization which secures the production and maintenance of the components that specify it in such a manner that the product of their functioning is the very same organization that produces them.”

The idea of autonomy was built on two main notions, as introduced by Piaget and further elaborated by the others. The first is *thermodynamic openness* (or openness to material causation, in Rosen’s vocabulary): an organism needs matter from the

environment in the form of building blocks from which to produce its components, and energy to perform the activities required to achieve self-production, self-repair, and self-maintenance and to interact with a changing environment. The second notion, which is distinctively biological, is *organizational closure* (or *closure to efficient causation*, in the case of Rosen): a biological organization is characterized as a closed network of processes of production in which each component is produced by others in the network such that the network maintains itself.

Departing from the traditional characterization of organizational closure and inspired by the work of Pattee (1972, 1973/2012) and Kauffman (2000), Moreno and Mossio (2015) emphasize the thermodynamics of organisms in a way that goes beyond the idea that organisms just need matter and energy—organisms must constrain free energy in constructing and maintaining their own components. The components that contribute to the construction and maintenance of a biological organism are characterized as constraints. These constraints canalize free energy into performing biological processes, including those responsible for the generation of other constraints. On this view, organisms must perform work to produce and maintain the very constraints that make the performance of work possible. The resulting account is one of *closure of constraints*: the existence and activity of the constraints operative in a living system depends on the action of other constraints in the system that direct the flow of free energy into their establishment.

Constraints can be organized in cycles: a constraint that enables one activity can be set by another simultaneous constraint, with each determining the other. However, the notion of closure involves a regress in which each constraint is constructed by the activity of one or more preexisting and already operative constraints until one arrives at the initial constitution of the organism. As a matter of fact, at birth some of these constraints are inherited from those produced by the parents (see Mossio & Pontarotti, 2022), but most of the constraints constituting an organism at any given moment have been produced, replaced, repaired and maintained by the organism during its lifetime.

The notion of closure of constraints fits well with the revisionist conception of mechanism explained in Sect. 5.2 as the appeal to constraints in Moreno and Mossio's account echoes the appeal to them in the revisionist account of mechanism. On both accounts, each activity of an organism is carried out through the constrained release of free energy. This is not surprising, as both accounts drew inspiration from Pattee. In this specific respect, one point of divergence between the two accounts concerns the entities responsible for those biological activities that, in the autonomy perspective, would coincide with biological functions as they contribute to the maintenance of the organism (Mossio et al., 2009). According to the account of closure of constraints, each of these functional activities is performed by one constraint, and closure consists in the mutual dependence between these functional constraints. To characterize these biological activities, the revisionist account of mechanism looks, instead, at organized sets of constraints, that is mechanisms, where the mechanisms are characterized by how they constrain the release of free energy.

As argued elsewhere (Bich & Bechtel, 2021), associating a single constraint with a biological function is an abstraction. While useful in some cases for explanatory purposes such as when considering an enzyme catalyzing a reaction, it risks overlooking the complexity underlying the realization of a biological function and how this complexity matters for the overall functioning of the system. Already in the relatively simple case of an enzyme, different parts, such as the catalytic site, the phosphorylation and allosteric sites, structures that undergo conformational changes, etc., contribute differently to the function performed by the enzyme. This function would be better characterized in terms of a mechanism employing several interacting constraints. This is even more evident in the cases of systems composed of components whose activity depends in turn on the interaction between different sub-components, such as in molecular complexes in cells. Likewise, in multicellular organisms, the activity of organs depends on the interaction of different structures (such as the muscles, valves, etc. in the heart) or cell types (for example alpha and beta cells, among others, in the pancreas) constituting them.

A possible way to connect closure of constraints and the revisionist account of mechanism is to consider functions as performed by mechanisms, which in turn are defined by their constraints. An interesting consequence of this conceptual step, which has plenty of implications to be explored in further work, is that biological mechanisms, and the constraints that they harbor, can be considered in the context of closure as dependent on the activities of other mechanisms (organized sets of constraints) in the organism.

At this point it is important to point out that this conceptual step is not as simple as it might seem at first sight and taking it would not be uncontroversial. New mechanism and autonomy are two complex frameworks which, however related and often intersecting or complementary, do not perfectly overlap, as they have different foci, strategies and different questions to which they aim to respond (Bich & Bechtel, 2022b). Closure of constraints differs from mechanistic accounts in that it emphasizes the relations between activities that contribute to the maintenance of the system, rather than between the component activities that mechanists treat as giving rise to phenomena. Moreover, it treats the organism as a whole as the starting point and the main focus when addressing what is distinctive about living organisms. It aims to identify what functions are necessary to produce and maintain it and how they depend on one another, rather than to explain how a specific biological phenomenon is materially realized. In doing so, work on autonomy does not engage in decomposition in the same way as described by the mechanists. Yet if one accepts that biological functions require mechanisms made of constraints rather than individual constraints, and considers the role of constraints in defining mechanisms, one might go as far in bringing the two frameworks together as to consider that organizational closure may be recharacterized as a special type of closure of mechanisms.

However, as we develop in the next sections, closure alone, although a fundamental notion, cannot account for the distinctive causal regime at work in biological systems. Control also plays a central role and needs to be taken into account.

5.4 Control Mechanisms

In discussing mechanisms in Sect. 5.2, we noted that the constraints that determine how a mechanism will behave can be influenced by other processes outside the mechanism. Such a process can itself be construed as due to the work of a different type of mechanism that constrains free energy to act on the constraints in the production mechanism. We refer to such mechanisms as *control mechanisms*. In order for control mechanisms to produce changes in production mechanisms that are appropriate to the circumstances within or confronting the organism, control mechanisms must be able to procure information about these circumstances. Following Pattee, we will characterize the process by which they do so as *measurement*. What this requires is that the constraints in control mechanisms that determine what action they perform be responsive to the circumstances within and confronting the organism. Many organisms rely on detecting chemicals in their environment and moving as a result. The chemicals alter constraints in the sensors and the altered constraints in the sensors result in changes in the production mechanism, altering what it does.

Allowing measurements to affect the constraints in control mechanisms seems to be in tension with the account of closure of constraints developed in Sect. 5.3. That required that each constraint, and hence each mechanism, be itself the product of work performed by other constraints constituting other mechanisms within the organism. But in order to make measurements, these constraints must be modified by what is being measured. Especially when control mechanisms make measurements of conditions external to the organism, this seems to undermine closure—a given constraint is causally modified by things other than the constraints constituting the organism.

The resolution to this challenge is to recognize that measurement is a different type of interaction from those involved in the production and maintenance of a constraint within a regime of closure. To see how closure of constraints can be maintained even as control mechanisms make measurements, we need to distinguish a constraint itself from the particular forms it may take. Consider a mechanical thermostat that controls a furnace by registering the temperature through a bimetallic coil. Higher temperatures cause the outer strips to expand more than the inner strip, resulting in the strip curving away from the point of contact that completes the circuit to the furnace, breaking the circuit. We can distinguish the constraints constituting the thermostat from the curvature of the strip at a given time. Both are involved in the action on the furnace, but the fixing of the constraints that constitute the thermostat—the constitution of the strip from two metals and the positioning of the contact point—is different from the fixing of the curvature on a particular occasion. The constitution of the thermostat determines what it can measure while the ambient air determines the actual measurement. A thermostat is designed to be informed by the temperature of the air and so is open to information. The same applies to control mechanisms within an organism such as a chemosensory neurons, except now the constitution of the measuring device is the product of the closed system of constraints constituting the organism. The constraints that enable the neuron to

measure the presence of a given chemical are established through the activity of other mechanisms within the organism while its actual registration on a given occasion carries information about the chemicals in the environment of the neuron. Control mechanisms are open to information even as the constraints that constitute them and enable them to do so are determined by other constraints with the organism, preserving closure.

Closure requires that each control mechanism operative in an organism be itself the product of another mechanism within the mechanism. Since causes must precede their effects, closure inevitably takes us back to the initial constitution of the organism. What is present at the beginning of the life of an organism is itself the product of another organism from which it was generated. Among other things, an organism begins life with both the mechanisms needed to recruit and constrain energy and its genetic material. While not acting on their own, genes play an important role in determining what further mechanisms the organism will construct, both production mechanisms and control mechanisms. Synthesis of new proteins involves transcription factors initiating the transcription of the sequence of nucleic acids constituting DNA into a corresponding sequence of nucleic acids in RNA, which is then, in the case of eukaryotic organisms, transported to the ribosomes in the cytoplasm, where it is translated into a corresponding sequence of amino acids. These are folded, often with assistance of other proteins, into proteins in the endoplasmic reticulum. Some of the newly minted proteins are prepared for export out of the cell in the Golgi apparatus, but others are incorporated into the structure of the cell, where they catalyze biochemical reactions. Genes thus provide a template for the proteins that subsequently perform the various activities cells carry out to maintain themselves. In providing such a template, genes quite literally inform (specify the constitution) of proteins. Genes have, accordingly, been viewed as constituting information. But there is a significant difference between the informational role of genes and that of external conditions measured by an organism's sensory systems. The information in genes determines (up to a certain degree) the constitution (the sequence of amino acids) of mechanisms, including control mechanisms. Other components of the new organism (acquired in part as a result of interacting with entities in the environment) determine which of these mechanisms will be constructed by determining which genes will be expressed.

At the outset and throughout life, genes specify the structure of the mechanisms constituting the organism. Transcription factors, and the mechanisms producing them, determine which genes will be expressed. Once produced, some of the resulting proteins constitute control mechanisms. Some control mechanisms determine subsequent gene expression and hence the constitution of subsequent mechanisms. These control mechanisms are informed not just by genes but by measurements the control mechanisms make. Measurements, as we discussed above, don't directly determine the constitution of the control mechanism but rather influence the form it takes, typically in a variable manner. However, over time they do end up contributing to the constitution of the organism by determining which genes are expressed (as well as what posttranslational modifications are made to the product proteins). Whereas at the outset all of the machinery constituting the organism originated in

the parent organism, over time the machinery reflects both its initial constitution and its experiences.

In this section we have developed the conception of control mechanisms as mechanisms that direct the productive activities of organisms while taking into account information reflecting the organism's condition and environment. They are the vehicle through which organisms determine how they will act to generate, repair, and maintain themselves. In virtue of each production and control mechanism being generated from other mechanisms constituting the organism, organisms manifest a closure of constraint even as they remain open to information and alter their behavior in light of this information.

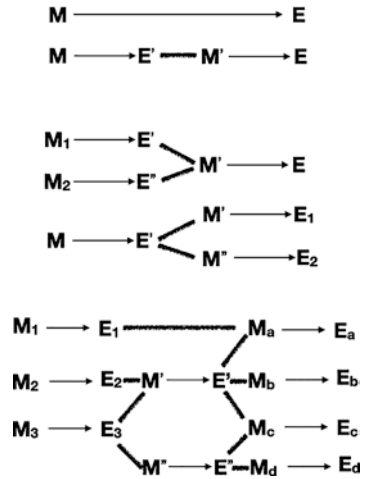
5.5 Integrating Control Mechanisms

Our characterization of control mechanisms identified two features—a measurement process that affects the constraints constituting the control mechanism and the action of those constraints on the constraints of other mechanisms (production or control mechanisms). The components of a control mechanism responsible for measurement and for acting on constraints in other mechanisms can be tightly coupled, as they are in a thermostat. But they can also be separated, with multiple components intervening between those carrying out the measurement and those acting on other mechanisms. Just as it is sometimes helpful to decompose production mechanisms into component production mechanisms, sometimes it is useful to decompose control mechanisms into component control mechanisms. For each of the component control mechanisms to satisfy our characterization of a control mechanism, each must make measurements and carry out action on other mechanisms. This can be accommodated if we view the connection between the two control mechanisms as involving signals—the generation of entities whose role is to be measured by another mechanism. Then one control mechanism can be viewed as generating the signal while the other can be viewed as measuring it by allowing its constraints to be informed by the signal (Fig. 5.1, top). These components can be separated by a distance but still work together in exercising control over a production mechanism.

Allowing for signals between control mechanisms greatly expands the potential for control. The same initial component that makes a measurement can generate multiple signals that are responded to by different control mechanisms, thereby allowing one measurement to effect control over multiple production mechanisms. Or the same downstream control mechanism can respond to signals arising from multiple control mechanisms and thereby respond to different measurements (Fig. 5.1, middle). These possibilities can be combined in various ways, resulting in multiple control processes interacting with each other (Fig. 5.1, bottom). The result might be viewed as a network of control mechanisms.

A network of control mechanisms is not just a theoretical possibility. It appears to be what exists in living organisms, including mammals. Even the simplest organism consists of multiple production mechanisms and individual production

Fig. 5.1 Complex interactions of control mechanisms. Top: a signal can mediate between two components of a control mechanism. Middle: one control mechanism can respond to signals from two different control mechanisms or one control mechanism can release a signal that is responded to by two different control mechanisms. Bottom: different control processes can be integrated into a network



mechanisms can be regulated by multiple control mechanisms, where these are connected by signals (Bich & Bechtel, 2022a). The multiplicity of control mechanisms raises the prospect that different control mechanisms will result in inconsistent actions, presenting challenges for the ability of the organism to maintain itself. How can multiple control mechanisms act to enable the organism to maintain itself?

One way to make individual components work together is to bring them under a single control mechanism that directs all of their activities. The framework we have developed allows for conceptualizing control in hierarchical terms. One control mechanism can operate on the constraints of multiple others (Fig. 5.1, middle). We can characterize the one operating on the others as at a higher level of control (Fig. 5.2a). This consolidation of control can be iterated over multiple levels with fewer controllers at each level in the hierarchy until there is just one at the top level. If the highest-level mechanism is appropriately constituted, it can impose directives on those below it so that, at the bottom of the hierarchy, production mechanisms operate in appropriate ways with respect to each other—e.g., different muscles contracting either simultaneously or in a specified sequence.

Hierarchical control is an intuitively attractive solution to insuring coherent operation of production mechanisms. It comes, however, with a significant cost—if it is to enable the organism to survive, the control mechanism at the top of the hierarchy must acquire all the information required to select appropriate actions. It must be constituted to make all the relevant measurements and, based on them, execute commands for all the appropriate actions. Such a hierarchy is compatible with lower-level control mechanisms procuring information appropriate to executing activities delegated to them. But if the organism is to maintain itself, the highest-level control must receive the information needed to determine the directives to send to control mechanisms subordinate to them in all the situations that the organism might confront. This would require an extremely sophisticated homunculus.

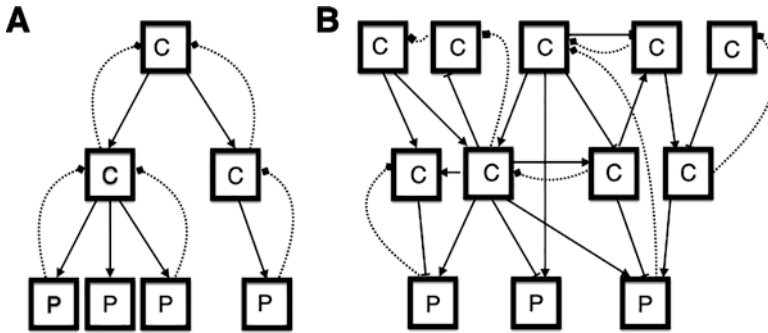


Fig. 5.2 (a) Hierarchical organization with information being transmitted to higher-level controllers (dotted lines) and control being executed (solid arrows) on production mechanisms or lower-level controllers. There are fewer controllers at each level, culminating in a single executive controller. (b) Heterarchical organization in which multiple controllers can operate in single production mechanisms. Although still presented in terms of levels, the occurrence of arrows directed horizontally and upwards indicates that the ranking of levels is breaking down. There is no single controller at the top. One might better characterize **b** as a network involving interactions that, in the case of control relationships, are only locally hierarchical

Hierarchy is not the only option. Human preferences not infrequently violate hierarchical preferences: an individual prefers A to B, B to C, and C to A. Yet people still function well in the world.⁸ For such non-hierarchical relations, McCulloch (1945) coined the term *heterarchy*. We can extend this concept to control mechanisms: it is possible for an organism to be so constituted that mechanism A controls mechanism B, B controls C, and C controls A. As with heterarchical preferences, heterarchical controllers may not be problematic: the different controllers might each respond to different information and may work together in different combinations to enable an organism to cope successfully with different environments. There is, moreover, no reason to restrict this scenario to three control mechanisms organized in a circle. Organisms consist of a multitude of control mechanisms, many of which act on other control mechanisms as well as production mechanisms. On this scenario, an organism consists of a network of control mechanisms that interact with each other in a multitude of ways (Fig. 5.2b). As long as each controller is a product of production mechanisms within the organism, one can have a highly dynamic network of controllers without violating closure.

Control mechanisms can be organized in a heterarchical manner that results in an organism responding to conditions it faces in ways that maintain itself. But the variety of heterarchical arrangements is immense and most heterarchical organizations of control processes are unlikely to result in an organism maintaining itself. Which types of heterarchical organization are likely to be successful? Rather than approaching this question a priori, we suggest drawing inspiration from biology. Control

⁸This can result in incoherent behavior, but it needn't. If one only confronts pairwise choices, then in each instance the relevant preference can yield a decision.

systems in current biological organisms have demonstrated success since they have succeed in keeping organisms earlier in the lineage alive. Because they are more familiar to most readers, we will focus on mammals.

Many of the control processes in mammals involve neurons, of which many are situated in the brain. Often neuroscientists focus their inquiries on the most recently evolved part of the brain—the neocortex. Within the neocortex, they often conceptualize the frontal regions of the neocortex as the central executive directing the activities of the organism. In doing so, researchers are implicitly assuming that neural control is organized hierarchically. But Sterling and Laughlin (2015) offer a contrary perspective, arguing that a principle exemplified by brains is local control. Another principle exemplified in brains, they argue, is to use chemistry whenever possible. This may seem surprising in an account that focuses on the brain since brains are often characterized as electrical processing systems. Neurons, however, carry out their control processes chemically. Neurons receive inputs from other neurons when neurotransmitters bind to their receptors and respond by performing chemical reactions (Bechtel, 2022). These may involve opening and closing ion channels, thereby affecting electrical currents across that cell's membranes. But in many cases, they carry out a variety of chemical reactions that alter the metabolism of the neuron. These activities include synthesizing new proteins. Moreover, many control activities of neural systems involve acting on the endocrine system, through which cells release molecules that travel through extracellular space (e.g., the blood stream) and act on other cells through receptors at their surface. The endocrine system is an important control system that often exercises control locally within tissues. Whether in the endocrine system or in the central nervous system, much control is carried out locally and chemically.

The role of control mechanisms is to enhance the probability that internal and external activities of the organism are performed when and in the way that is needed for and compatible with the maintenance of the organism itself (Bich et al., 2016). However, another primary role of control in keeping organisms alive is to maintain production mechanisms in conditions in which they can operate. The importance of this was emphasized by one of the first biologists to emphasize the control of biological mechanisms—Claude Bernard. Bernard (1878) described each production mechanism as operating to maintain the fixity or constancy of what he termed the *internal environment*. For Bernard, the result was to free birds and mammals from the vicissitudes of the external environment: whenever factors in the external environment perturbed conditions within the organism, one or more mechanism would be activated to perform its activity to restore the internal environment. As a result, each mechanism could rely on a stable internal environment and was free from the vicissitudes of the external environment.

Bernard did not describe the processes whereby such control was executed. This endeavor was taken up by Cannon (1929, 1932), who introduced the notion of *homeostasis* to characterize the processes through which organisms maintain themselves in similar conditions. In particular, he pays specific attention to the maintenance of some of the features of the “fluid matrix of the body” (citing Bernard's characterization of the *internal environment* as the “totality of the circulating fluids

of the organism,” Cannon, 1932 p. 38). This matrix includes blood and lymph, and some of its features to be maintained are temperature, pressure, and concentrations of ions and molecules. Although Cannon described other means of maintaining homeostasis such as buffering, his main examples involved negative feedback. By this time human designers had identified negative feedback as an effective means of maintaining mechanical systems in a constant target state. Subsequently negative feedback was adopted by the cyberneticists as providing a primary means for controlling biological, social, and engineered systems (Wiener, 1948). For many, homeostasis became identified with negative feedback.

Negative feedback provides a useful starting place for understanding how local control can help maintain the organism. Negative feedback involves measuring a product produced by a production mechanism and, if the value falls outside a target range, acting on one or more constraints in the producing mechanism to alter its function. For example, if pancreatic β cells detect that glucose levels in the blood exceeds a target, they increase the synthesis of insulin and release massive amounts of it into the blood, where it can bind receptors on different cell types. When this high amount of insulin binds to receptors on liver cells, it speeds up glucose intake and the process in which glucose is converted to glycogen, thereby reducing the concentration of glucose in the blood. This process stops when blood glucose concentrations drop below the level that stimulates high insulin release.

In many circumstances, local negative feedback control of individual production mechanisms can provide a relatively constant environment. But it has its limits. To the degree that an organism has stored glycogen, negative feedback can restore glucose levels when they drop too low. This is achieved by releasing glucagon from pancreatic α cells, which stimulates gluconeogenesis from glycogen. But over time the supply of glycogen will be exhausted, and the organism must procure additional nutrients if it is to maintain sufficient glucose levels to fuel the organism’s production and control mechanisms. This requires control processes that initiate other activities such as those involved in feeding. For this, mammals rely on other hormones, for example ghrelin and leptin, being transported to the arcuate nucleus of the hypothalamus, a location in the brain without a blood-brain barrier at which hormones can act on the receptors of neurons. Ghrelin signals lack of food in the digestive system while leptin signals presence of fat. By measuring these and other physiological states of the organism and integrating them, neurons in the arcuate nucleus detect the need for eating and signal to neurons elsewhere in the hypothalamus act to initiate feeding activities.

The hypothalamus consists of multiple nuclei each comprising different populations of neurons, many of which respond to endocrines and are involved in releasing endocrines as well as neurotransmitters. Moreover, they often signal to each other with peptides or neuroendocrines, which are also distributed through the extracellular matrix. Some of these neurons, such as those that respond to ghrelin and leptin in the arcuate nucleus, are specialized for one type of activity (registering hunger or satiety in the case of agouti-related protein expressing neurons and pro-opiomelanocortin expressing neurons respectively). But cells in other nuclei, such as the lateral hypothalamus (one of the sites to which neurons in the arcuate nucleus

signal) receive multiple signals and send multiple outputs. The orexin neurons provide an illustrative example. They were so named after the Greek word for appetite since they were first identified as promoting eating activities (Sakurai et al., 1998). But they were subsequently implicated in an animal transitioning from sleep to waking (Adamantidis et al., 2007). Tsunematsu et al. (2013) showed that silencing them sufficed to induce slow wave sleep. Orexin neurons illustrate a common theme exemplified by many nuclei in the hypothalamus and other brain regions—they integrate signals from multiple sources and send signals (chemical and electrical) to multiple other centers, some leading to action (Fig. 5.3). The result is that control mechanisms regulating individual production mechanisms are coupled together so that information procured to control one production mechanism is also employed to control other production mechanisms. Accordingly, control of individual production mechanisms takes into account a wide range of conditions in the organism. This appears to be a mode of heterarchical organization that is effective in enabling organisms to maintain themselves.

By starting with negative feedback, we have treated control as a reactive process—each negative feedback control mechanism begins with measuring conditions and responding to that information. Even when these are integrated, the process starts with measuring a condition in the organism or its environment. But control mechanisms are capable of anticipatory control as well: they can enable an

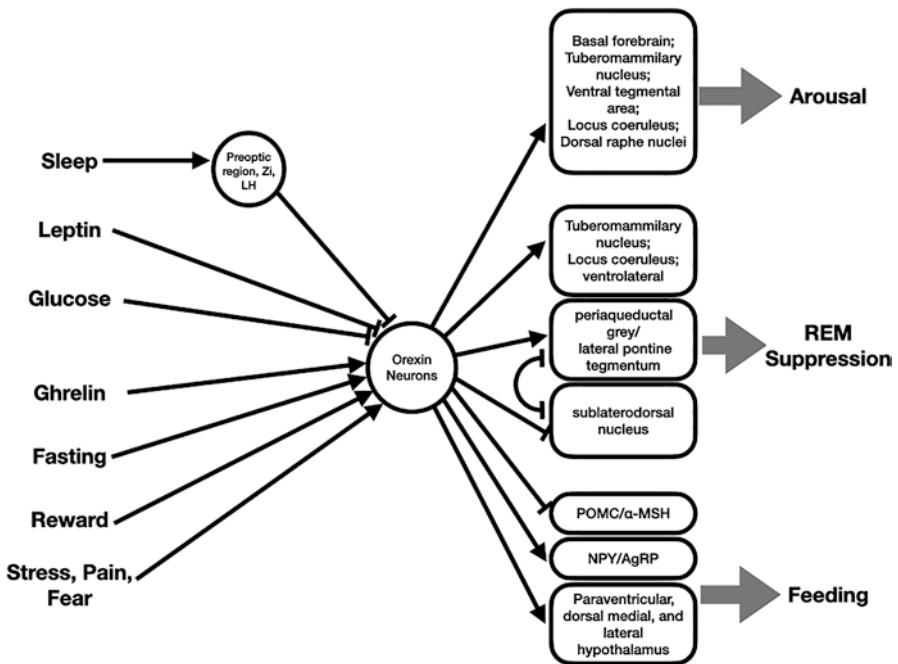


Fig. 5.3 Orexin neurons in the lateral hypothalamus respond to signals released by multiple other control mechanisms and have effects on multiple behaviors. (Based on data in Arrigoni et al. (2019))

organism to regulate its production mechanisms so that they operate in ways appropriate for conditions the organism is likely to confront in the future. For example, an organism's environment regularly presents different conditions at different times of day and, except in the tropics, during different seasons of the year. Controlling production mechanisms in ways that anticipate these conditions is facilitated by another nucleus in the hypothalamus, the suprachiasmatic nucleus (SCN), which generates oscillations with a period of approximately 24 hours (hence, circadian). Both through electrical signaling and release of peptides, neurons in the SCN send signals that are responded to either directly by production mechanisms or by neurons controlling other production mechanisms. The circadian system enables organisms to anticipate events that have occurred in a regular fashion over the phylogeny of the organism (Moore-Ede, 1986). Associative learning, achieved by changing constraints in neuroreceptors, provides a means to modulate control activities in light of regularities experienced by an organism in its lifetime. By modifying the constraints within neurons that determine how they integrate information to control various activities, neurons in nuclei in the hypothalamus and other brain regions enable organisms to initiate activities appropriate to events that are likely to follow. Accordingly, control mechanisms can be both reactive and anticipatory.

What this brief consideration of the hypothalamus suggests is that the control mechanisms that enable organisms to maintain themselves are often specific to the production mechanism being controlled. When coordination of control mechanisms is required, different control mechanisms interact with each other so that measurements procured in the control of one activity can also modulate the control of other activities. Control mechanisms do operate on other control mechanisms, but rather than assuming control over subordinates, these control mechanisms integrate multiple measurements and, based on the result, modify constraints in the more local control mechanisms. Often the control needed to maintain the organism involves locomotor activity that procures food or avoids dangerous situations. There is not space to develop the account here, but this involves the control of skeletal muscles. Here too control is primarily specific to the production mechanism. Individual muscles are controlled by pattern generators which are then coupled to enable multiple muscles to coordinate their contraction. The activity of these pattern generators can be modulated by signals not only from other pattern generators but also by those from neurons in different nuclei of the hypothalamus or other brain regions that register conditions requiring behavioral adjustment.

Of particular significance in coordinating different production mechanisms are the so called neuromodulators—transmitters such as dopamine, serotonin, and numerous neuropeptides. In response to measured conditions in the organism, these neurotransmitters are released into extracellular space and diffuse to neurons with appropriate receptors. They act on a relatively long timescale (e.g., seconds and minutes), transforming the context in which other neural processing occurs. Due to the extended space and time in which they act, they can modulate the behavior of many specific controllers (Katz, 1999). Despite their importance in directing overall activity both in the brain and the organism, neuromodulators do not instantiate a hierarchical system. Each is released by different nuclei in the brain and promotes

different activities. The control they execute is heterarchical (Bechtel, 2022). When there are conflicts, the determination as to which activities to carry out is made by another set of nuclei, those of the basal ganglia. These nuclei enable a competition between control mechanisms, inhibiting all but the winner of the competition, thereby avoiding conflicts between them (Bogacz & Gurney, 2007). However, the basal ganglia are not themselves in control—they are simply another component in a heterarchically organized network of control mechanisms (Bechtel & Huang, 2020).

Taking our cue from mammals, we see that the control mechanisms that serve to maintain organisms are organized heterarchically, not hierarchically, with much control remaining specific to the production mechanisms being controlled. When it is important to coordinate multiple responses, control mechanisms are employed that integrate multiple control mechanisms. Of course in mammals these control processes are supplemented with other control mechanisms such as those in the neocortex. The distinctive potency of neocortical processing is exhibited in visual processing. Whereas many vertebrates rely primarily on the tectum/superior colliculus to coordinate visually acquired information directly with motor activity, by relying on the neocortex, higher mammals can engage in more complex categorization and learning in response to visual inputs. But, as illustrated by the ability of decerebrate cats to live on their own, albeit in protected environments (Björsten et al., 1976), cortical processing is not required for many of the activities organisms perform in the service of their self-maintenance. Moreover, when processing is carried out in the neocortex, it must be coupled with the more basic control mechanisms on which we have focused in order to affect behavior. Sub-cortical control mechanisms are fundamental to the ability of organisms, including humans, to maintain themselves.

5.6 Conclusions

Organisms need mechanisms to construct, repair, and maintain themselves. A major difference between human-made machines and biological mechanisms is that biological mechanisms are dependent on the organism of which they are part to construct, maintain, and repair them. Organisms and biological mechanisms are mutually dependent: without the organism, biological mechanisms wouldn't exist and endure; without mechanisms, the organism would not maintain itself. Our contention has been that this mutual dependence is mediated by control mechanisms. Without control mechanisms, production mechanisms will simply carry out their activities any time what Machamer et al. call start or set-up conditions are satisfied. They won't tailor their activities to what the organism needs to maintain itself. Only if production mechanisms are controlled will they perform their activities when and in the manner needed to maintain the organism.

Our discussion of control mechanisms reveals two complementary features. On the one hand, the constraints constituting control mechanisms are the product of the

mechanisms that constitute the organism. On the other hand, the particular values they take are determined by the measurements they make. By measuring appropriate variables, control mechanisms are able to act on production mechanisms so that they serve the needs of the organism. This very ability, though, is determined by how they are constituted by other mechanisms in the organism. They are thereby part of the closure of constraints but also open to the information that is relevant to whether the actions of production mechanisms are needed and useful to the organism.

We have emphasized that organisms exhibit a multitude of control mechanisms. If they are the basis for organisms successfully maintaining themselves, their activity needs to be coordinated. Although hierarchical organization would ensure coherence, we have argued that in biological systems control is organized heterarchically. Inspired by biology, we suggest that effective heterarchical control involves controllers of specific mechanisms being integrated into networks in which information procured by different control mechanisms is shared and used to constrain the behavior of the different control components. Such heterarchical networks, crafted over the course of evolution, appear to be what enable organisms to maintain themselves while engaging dynamic environments.

Funding The authors acknowledge funding from the Basque Government (Project: IT1228-19 and IT1668-22 for LB), Ministerio de Ciencia, Innovación y Universidades, Spain (research project PID2019-104576GB-I00 for LB and WB, and ‘Ramon y Cajal’ Programme RYC-2016-19798 for LB) and the John Templeton Foundation (Project 62220 for LB).

References

- Abrahamsen, A., & Bechtel, W. (2011). From reactive to endogenously active dynamical conceptions of the brain. In K. Plaisance & T. Reydon (Eds.), *Philosophy of behavioral biology* (pp. 329–366). Springer.
- Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K., & de Lecea, L. (2007). Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature*, *450*(7168), 420–424. <https://doi.org/10.1038/nature06310>
- Arrigoni, E., Chee, M. J. S., & Fuller, P. M. (2019). To eat or to sleep: That is a lateral hypothalamic question. *Neuropharmacology*, *154*, 34–49. <https://doi.org/10.1016/j.neuropharm.2018.11.017>
- Bechtel, W. (2022). Reductionistic explanations of cognitive information processing: Bottoming out in neurochemistry. *Frontiers in Integrative Neuroscience*, *16*, 944303.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *36*(2), 421–441.
- Bechtel, W., & Abrahamsen, A. (2009). Complex biological mechanisms: Cyclic, oscillatory, and autonomous. In C. A. Hooker (Ed.), *Philosophy of complex systems. Handbook of the philosophy of science* (Vol. 10, pp. 257–285). Elsevier.
- Bechtel, W., & Bollhagen, A. (2021). Active biological mechanisms: Transforming energy into motion in molecular motors. *Synthese*, *199*(5–6), 12705–12729. <https://doi.org/10.1007/s11229-021-03350-x>
- Bechtel, W., & Huang, L. T. (2020). Decentering cognition. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd annual meeting of the cognitive science society* (pp. 3247–3253). The Cognitive Science Society.

- Bechtel, W., & Richardson, R. C. (1993/2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. MIT Press. 1993 edition published by Princeton University Press.
- Bernard, C. (1878). *Leçons sur les phénomènes de la vie communs aux animaux et aux végétaux*. Baillière.
- Bich, L., & Bechtel, W. (2021). Mechanism, autonomy, and biological explanation. *Biology and Philosophy*, 36(6). <https://doi.org/10.1007/s10539-021-09829-8>
- Bich, L., & Bechtel, W. (2022a). Control mechanisms: Explaining the integration and versatility of biological organisms. *Adaptive Behavior*, 30(5), 389–407. <https://doi.org/10.1177/10597123221074429>
- Bich, L., & Bechtel, W. (2022b). Organization needs organization: Understanding integrated control in living organisms. *Studies in History and Philosophy of Science*, 93, 96–106. <https://doi.org/10.1016/j.shpsa.2022.03.005>
- Bich, L., Mossio, M., Ruiz-Mirazo, K., & Moreno, A. (2016). Biological regulation: Controlling the system from within. *Biology and Philosophy*, 31(2), 237–265. <https://doi.org/10.1007/s10539-015-9497-8>
- Bjursten, L. M., Norrsell, K., & Norrsell, U. (1976). Behavioural repertory of cats without cerebral cortex from infancy. *Experimental Brain Research*, 25(2), 115–130. <https://doi.org/10.1007/BF00234897>
- Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, 19(2), 442–477. <https://doi.org/10.1162/neco.2007.19.2.442>
- Bollhagen, A., & Bechtel, W. (2022). Discovering autoinhibition as a design principle for the control of biological mechanisms. *Studies in History and Philosophy of Science*, 95, 145–157. <https://doi.org/10.1016/j.shpsa.2022.08.008>
- Cannon, W. B. (1929). Organization for physiological homeostasis. *Physiological Reviews*, 9, 399–431.
- Cannon, W. B. (1932). *The wisdom of the body*. W. W. Norton & Company, inc.
- Fisher, A. J., Smith, C. A., Thoden, J., Smith, R., Sutoh, K., Holden, H. M., & Rayment, I. (1995). Structural studies of myosin-nucleotide complexes: A revised model for the molecular basis of muscle contraction. *Biophysical Journal*, 68(4), S19–S28.
- Fiske, C. H., & Subbarow, Y. (1929). Phosphorus compounds of muscle and liver. *Science*, 70, 381–382.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 50–71.
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.
- Holmes, K. C., & Geeves, M. A. (2000). The structural basis of muscle contraction. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 355(1396), 419–431. <https://doi.org/10.1098/rstb.2000.0583>
- Hooker, C. A. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science*, 18(4), 757–780. <https://doi.org/10.1007/s10699-012-9304-9>
- Huxley, H. E. (1969). The mechanism of muscular contraction. *Science*, 164(3886), 1356–1365.
- Kant, I. (1790/1987). *Critique of judgement*. Hackett Publishing.
- Katz, P. S. (Ed.). (1999). *Beyond neurotransmission: Neuromodulation and its importance for information processing*. Oxford University Press.
- Kauffman, S. A. (2000). *Investigations*. Oxford University Press.
- Lavoisier, A. L., & Laplace, P. S. d. (1780). Mémoire sur la Chaleur. *Mémoires de l'Académie royale des sciences*, 35–408.
- Liebig, J. (1840). *Organic chemistry in its applications to agriculture and physiology*. Taylor and Walton.
- Lohmann, K. (1929). Über die Pyrophosphatfraktion im Muskel. *Naturwissenschaften*, 17, 624–625.
- Lynn, R. W., & Taylor, E. W. (1971). Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry*, 10(25), 4617–4624. <https://doi.org/10.1021/bi00801a004>

- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25. <https://doi.org/10.1086/392759>
- Maturana, H. R. (1980). Biology of cognition. In H. R. Maturana & F. J. Varela (Eds.), *Autopoiesis and cognition: The realization of the living* (pp. 1–58). Reidel.
- Maturana, H. R., & Varela, F. J. (1980). Autopoiesis: The organization of the living. In H. R. Maturana & F. J. Varela (Eds.), *Autopoiesis and cognition: The realization of the living* (pp. 73–138). Reidel.
- McCulloch, W. S. (1945). A heterarchy of values determined by the topology of nervous nets. *The Bulletin of Mathematical Biophysics*, 7(2), 89–93. <https://doi.org/10.1007/BF02478457>
- Mendelsohn, E. (1964). *Heat and life: The development of the theory of animal heat*. Harvard University Press.
- Moore-Ede, M. C. (1986). Physiology of the circadian timing system: Predictive versus reactive homeostasis. *The American Journal of Physiology*, 250(5 Pt 2), R737–R752. <https://doi.org/10.1152/ajpregu.1986.250.5.R737>
- Moreno, A., & Mossio, M. (2015). *Biological autonomy: A philosophical and theoretical inquiry*. Springer.
- Mossio, M., & Pontarotti, G. (2022). Conserving functions across generations: Heredity in light of biological organization. *The British Journal for the Philosophy of Science*, 73(1), 249–278. <https://doi.org/10.1093/bjps/axz031>
- Mossio, M., Saborido, C., & Moreno, A. (2009). An organizational account of biological functions. *The British Journal for the Philosophy of Science*, 60(4), 813–841. <https://doi.org/10.1093/bjps/axp036>
- Mossio, M., Bich, L., & Moreno, A. (2013). Emergence, closure and inter-level causation in biological systems. *Erkenntnis*, 78(2), 153–178. <https://doi.org/10.1007/s10670-013-9507-7>
- Nicholson, D. J., & Dupré, J. (Eds.). (2018). *Everything flows*. Oxford University Press.
- Pattee, H. H. (1972). Laws and constraints, symbols and languages. In C. H. Waddington (Ed.), *Towards a theoretical biology*. Adine-Atherton.
- Pattee, H. H. (1973/2012). The physical basis and origin of hierarchical control. In *Laws, language and life* (Vol. 7, pp. 91–110). Springer Netherlands.
- Piaget, J. (1967). *Biologie et connaissance*. Gallimard.
- Rosen, R. (1972). Some relational cell models: The metabolism-repair systems. In R. Rosen (Ed.), *Foundations of mathematical biology* (Vol. II, pp. 217–253). Academic.
- Rosen, R. (1991). *Life itself: A comprehensive inquiry into the nature, origin, and fabrication of life*. Columbia University Press.
- Ryle, G. (1949). *The concept of mind*. Barnes and Noble.
- Sakurai, T., Amemiya, A., Ishii, M., Matsuzaki, I., Chemelli, R. M., Tanaka, H., et al. (1998). Orexins and orexin receptors: A family of hypothalamic neuropeptides and G protein-coupled receptors that regulate feeding behavior. *Cell*, 92(4), 573–585.
- Sterling, P., & Laughlin, S. (2015). *Principles of neural design*. MIT Press.
- Swan, J. A., Sandate, C. R., Chavan, A. G., Freeberg, A. M., Etwaru, D., Ernst, D. C., ... Partch, C. L. (2021). Hidden conformations differentiate day and night in a circadian pacemaker. *bioRxiv*, 2021.2009.2014.460370. <https://doi.org/10.1101/2021.09.14.460370>
- Tsunematsu, T., Tabuchi, S., Tanaka, K. F., Boyden, E. S., Tominaga, M., & Yamanaka, A. (2013). Long-lasting silencing of orexin/hypocretin neurons using archaerhodopsin induces slow-wave sleep in mice. *Behavioural Brain Research*, 255, 64–74. <https://doi.org/10.1016/j.bbr.2013.05.021>
- Wiener, N. (1948). *Cybernetics: Or, control and communication in the animal and the machine*. Wiley.
- Winning, J., & Bechtel, W. (2018). Rethinking causality in neural mechanisms: Constraints and control. *Minds and Machines*, 28(2), 287–310. <https://doi.org/10.1007/s11023-018-9458-5>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 6

Searching for Protein Folding Mechanisms: On the Insoluble Contrast Between Thermodynamic and Kinetic Explanatory Approaches



Gabriel Vallejos-Bacelliere and Davide Vecchi

Abstract The protein folding problem is one of the foundational problems of biochemistry and it is still considered unsolved. It basically consists of two main questions: what are the factors determining the stability of the protein's native structure and how does the protein acquire it starting from an unfolded state. Since its first formulation, two main explanatory approaches have dominated the field of protein folding research: a thermodynamic approach focused on energetic features and a kinetic approach focused on the temporal development of protein chains and structural considerations. Although these two approaches are tightly intertwined in biochemical practice and largely agree on which are the parts and activities in which the phenomenon under study should be decomposed to, there nevertheless exist important contrasts that have had repercussions on the development of the field and still engender vigorous debate. We shall analyse the historical development of the field and crucial aspects of current scientific debates. On this basis, we argue that the main sources of disagreement centre on the causal interpretation of thermodynamic and kinetic explanations, on the explanatory relevance assigned to different features of the phenomena under study and on the status of the ontological assumptions concerning the entities under study.

Keywords Protein folding problem · Thermodynamic hypothesis · Levinthal's paradox · Energy landscape · Brownian motion · Foldons · Equilibrium explanation

G. Vallejos-Bacelliere (✉)

Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Santiago, Chile

e-mail: gvallejos@ug.uchile.cl

D. Vecchi

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_6

109

6.1 Introduction: What Is the Protein Folding Problem

Proteins are central components in the functioning of all known organisms, carrying out functions such as catalysis, regulation of cell processes, transport, movement (from the subcellular to the organismal level), signalling, body construction etc. Without much exaggeration, proteins are what make life as we know it possible.

Proteins are linear polymers composed of amino acids linked together by peptide bonds.¹ Polypeptide chains are synthesized in the cell by the ribosomes, which catalyse the formation of the peptide bonds between its different amino acids, which are thus arranged in a polypeptide in accordance with the order of the nucleotide sequence of a messenger RNA (mRNA). The sequences of both polymers are related through the “genetic code”, which maps each amino acid (carried by a tRNA molecule with a specific “anticodon”) to a specific triplet (“codon”) of nucleotides of the mRNA sequence.² In an analogous way, the sequence of the mRNA is generated, in accordance with the principle of complementarity, by an RNA polymerase according to the order of the sequence of nucleotides in a DNA segment, i.e., a gene. Polypeptide chains are synthesized as random coils. However, in order to realize their function and being soluble,³ a protein must get folded into a specific compact 3D structure (or some restricted set of 3D conformations), i.e., what is called its “native state” or “native structure”,⁴ which is stabilized by different kinds of interactions between its components, like hydrogen bonds, van der Waals (VDW) interactions, electrostatic effects, hydrophobic effects, etc.⁵ The developmental process by which a protein undergoes a series of compositional and structural changes to acquire this final structure is called “folding”.

The protein folding problem (PFP) is one of the foundational problems of biochemistry. Since its formulation in the early 50s, it has spurred a substantial amount

¹A peptide bond is a covalent bond formed between the amino group of an amino acid and the carboxylic group of another. Because this bonded structure forms the backbone of the protein, proteins are also called polypeptides or polypeptide chains.

²The ‘genetic code’ is not strictly universal (Krebs et al., 2018).

³Usually, an unfolded protein is insoluble. The accumulation of insoluble components inside a cell beyond a certain threshold would be deleterious, causing stress or even tissue damage (Austin, 2009).

⁴It is usual for biochemists to use the term “3D-structure” to refer to specific conformations with low degrees of freedom that are stabilized by non-covalent interactions between their components. The native state is one of them. In this article, we centre our attention on globular proteins, i.e., proteins whose native state is a compact and water-soluble spherical-like conformation. This focus is justified by the fact that globular proteins are the most studied in protein folding research. Thus, research on their folding can be considered a research field on its own.

⁵This stabilization process might also involve environmental or non-intrinsic factors, that is, extrinsic vis-à-vis the intrinsic properties of the components of the developing protein, e.g., water, protons, ions, cofactors, prosthetic groups, ligands or other proteins (which can be other polypeptide chains of the same type in the case of homooligomers). Entropic factors may also play relevant causal roles, like the increase in solvent entropy caused by the burial of the hydrophobic moieties of the macromolecule. See Santos et al. (2020) for an analysis of the causal role of extrinsic factors in protein folding.

of theoretical and experimental research. Basically, the problem consists in answering the questions concerning what factors determine the stability of the native structure and how polypeptide chains reach their final native structure in a given medium.⁶ The problem is far from being trivial. Proteins are compositionally and structurally very complex entities. Given the high number of degrees of freedom of a polypeptide chain, which depend on the vast number of possible 3D arrangements of the component parts, a protein can, in principle, acquire an enormous number of possible conformations. However, despite this complexity, proteins fold rapidly, which means that from all the possible conformations, only a very restricted set is selected, leading to stable, soluble and functional 3D structures that are generally acquired in less than a minute (or less than a second in the case of smaller proteins).⁷

The protein folding problem has important consequences not only for basic biochemical research, but also in applied fields such as biomedicine and industry. Indeed, the aetiology of many degenerative diseases, such as Alzheimer and Creutzfeldt-Jakob disease, are related to the misfolding of proteins and the formation of insoluble protein aggregates that seem to destroy neuronal tissues (Liu et al., 2019). Moreover, the possibility of predicting the native state of proteins – given knowledge of the polypeptide chain – would provide significant understanding concerning the causal role of each gene in the case of any kind of organism. This predictive accomplishment would also imply the identification of new possible pharmacological targets for the cure of different kinds of medical conditions. For industry, knowing the factors that stabilize proteins would allow to develop technological applications, for example, the production of more stable enzymes.

⁶It is important to distinguish the protein folding problem, which is addressed in this article, from the protein structure prediction problem. The latter corresponds to the aim of predicting the native structure of a protein given its amino acid sequence. In other words, the first is concerned with explanatory aims, while the second with predictive ones. In the beginning of protein folding research, both explanatory and predictive aims were indissociable. However, with the emergence of structural databases, it became possible to make predictions based only on observed structural patterns irrespective of the physical basis of its relative frequencies (i.e., prediction became independent from explanation, see Vallejos-Baccelliere, 2022). In this latter context, revolutionary advances have emerged thanks to the application of powerful computational methods and artificial intelligence, being softwares like AlphaFold2 and RoseTTAFold prominent examples of this field. In this article, when referring to predictive aims, it will only be in the context of its association to the explanatory aims of protein folding research.

⁷It is debatable whether intrinsically disordered proteins (IDPs) are an exception in this sense. On the one hand, while IDPs are characterised as a class of proteins that do not get folded to perform their function (i.e., they do not possess an identifiable native structure), many IDPs acquire some kind of “native” structure when performing their function (Gomes & Faísca, 2019), e.g., acquiring some form of “order” in a specific region as, for example, when interacting with some ligand. On the other hand, it is currently accepted that almost all proteins have regions with some degree of “disorder” (Medina et al., 2021).

6.2 Brief Historical Overview of Folding Research

For many years, the problem of how proteins acquire their native structure remained elusive. Because all proteins were extracted already folded from living cells, it was thought that proteins must get necessarily folded by some part of the cell machinery acting as a structural template. Given the high variety of protein types in a cell, this machinery was believed to have templates for each one (Tanford & Reynolds, 2003).⁸ Due to their ubiquity in organisms and their importance in almost all cell functions, an obvious proposal was to consider this machinery as constituted by proteins. However, this proposal generates a conundrum: if the templates necessary for folding any protein are other proteins, then how are such templates folded? Any answer to this puzzle seems to lead to an infinite regress. Later, in the context of protein synthesis research, it was proposed that ribosomes must carry out this template function. Nonetheless, all ribosomes in a cell turned out to have very similar structure, so the problem of how it is possible to have a template for all the variety of proteins remained unsolved (Tanford & Reynolds, 2003).

The problem was completely reformulated at the start of the 1960s thanks to the work of Christian Anfinsen and his collaborators (Anfinsen et al., 1961; Anfinsen, 1973).⁹ In a series of experiments, they managed to show that a purified protein (i.e., bovine pancreatic ribonuclease A), after being unfolded¹⁰ (i.e., losing its folded structure without the breaking of its peptide bonds) with urea and reducing agents, could be reversibly refolded once the denaturants were extracted from the medium, thus recovering its biological activity in the absence of any other cellular component. Based on these results, Anfinsen proposed the so called “thermodynamic hypothesis”: “... *the three-dimensional structure of a native protein in its normal physiological milieu is the one in which the Gibbs free energy of the whole system is lowest; that is, the native conformation is determined by the totality of the interatomic interactions and hence by the amino acid sequence, in a given environment*” (Anfinsen, 1973, p. 223). The hypothesis seemingly makes two different kinds of claim: the first is that the native structure of a protein corresponds to the conformation with minimal free energy; the second, of implicit causal nature, is that

⁸The original postulation of templates was speculative, referring to an entity that actively folds proteins by performing the role of master mould to be copied. This hypothesis was promoted, among others, by the defenders of the ‘one gene, one enzyme’ hypothesis (Tanford & Reynolds, 2003). This putative template role is different from that of entities either accelerating folding or restricting the possible conformations acquired by the polypeptide. For instance, chaperones, instead of acting as templates, perform other functions, such as the acceleration of the folding process, the isolation of the folding protein from the external environment (i.e., “Anfinsen’s cage”) or the partial restriction of the possible alternative conformations (Sorokina et al., 2022).

⁹Anfinsen received the chemistry Nobel Prize in 1972 for this work.

¹⁰Denatured and unfolded are different concepts. In the present context, the first is functional, referring to loss of biochemical activity (e.g., the loss of the catalytic capacity of RNase A). The second concept is structural, referring to the loss of 3D structure of a protein.

the native structure is determined by the amino acid sequence. Let us analyse each claim in turn and, particularly, their relation.¹¹

The first kind of claim has been interpreted as meaning that, of all the possible conformations a protein might acquire, the native structure is the most stable in the appropriate “physiological conditions”. This assertion is both independent of any consideration concerning the temporal development of a polypeptide chain from the unfolded to the native state and also independent of the possible regulation of the folding process *in-vivo*. The corollary of this view is that the folding process is “spontaneous”¹² in the specific sense that it is merely thermodynamically driven. In microscopic terms, this hypothesis is currently represented by a conformational energy landscape with a funnel-like shape characterised by a single global minimum.¹³ The total free energy of each possible conformation would be determined by the contribution of the totality of interatomic interactions that are established in each token case. Therefore, these interactions are what determine the shape of a protein’s energy landscape.

The second claim is of a causal nature and asserts that the native structure of a protein in a given medium is “determined” by its amino acid sequence. This claim accounts for the observed refolding capacities of the, by supposition, completely unfolded protein¹⁴ when denaturants are extracted from the medium in the absence of other cellular components. The spontaneity of folding is then explained by the intrinsic properties and potentialities possessed by the amino acid components of the polypeptide chain, which are immutable and unaffected by any causal interaction of the developing protein with extrinsic factors (Santos et al., 2020). Then, the acquisition of the native structure by a protein, either during translation or when folding/refolding, is due to the “activation” (or “manifestation”) or “suppression” of these potentialities. The role of the environment, which includes the interaction of the developing protein with extrinsic factors and regulatory processes – that, *in vivo*, include the causal role of cellular components (ribosomes, other proteins, chaperones, ligands, etc.) coupled to other energetic processes (e.g., ATP hydrolysis) -, would then only be that of activating or suppressing some of the immutable

¹¹ For a discussion of the dual nature of the thermodynamic hypothesis, see Santos et al., 2020.

¹² In thermodynamics, a process is defined as spontaneous when the initial state has a higher free energy than the final one. This definition is independent of how the transformation between states occurs and, hence, it is also independent of the time that this transformation would take.

¹³ Revisions of the thermodynamic hypothesis taking into consideration a plurality of local equilibria have of course been proposed (Dill & Chan, 1997). Such revisions are necessary to account for the existence of, for instance, conformational changes, folding intermediaries, misfolding and amyloid fibril structures (whereby the latter are more stable than the native one). This revision requires the postulation of shallower minima of the energy landscape. However, any pluralistic move seems to imply a weakening of Anfinsen’s original hypothesis; this is because the hypothesis that folding is a merely thermodynamically driven process (instead of being directed or regulated) becomes increasingly questionable the more complex the shapes of the energy landscape are.

¹⁴ In biochemistry, a frequently discussed issue is to characterize what a denatured state really is and whether there really are “completely unfolded” proteins (see Sorokina et al., 2022 for a good synthesis).

potentialities, a process that, in a physiological medium, will lead to the formation of the native state. Under this interpretation, phenomena like misfolding or the denaturation, aggregation and degradation through time (which occur in the natural cellular milieu and in-vitro) is attributable to "...secondary effects [that] are claimed to shift the equilibrium towards the unfolded state, preventing thermodynamically-driven folding" (Sorokina et al., 2022, p. 7), i.e., preventing the manifestation of the aforementioned intrinsic potentialities.

The thermodynamic hypothesis opened a whole new field of research whose aim was to predict the native state of each protein with known sequence by seeking its minimal energy conformation among the totality of possible ones. This research programme relied on already accepted knowledge about the physical properties common to all molecules, which would allow to obtain each conformation of a macromolecule just by tinkering with it.¹⁵ This body of knowledge concerned, for instance, the nature of the covalent bond and molecular geometries, including the length of the covalent bonds, the possible angles between bonds, the rotatability of two moieties separated by a single bond, the planarity of a double bond etc. Given this knowledge, it would become in principle possible to describe all the possible steric restrictions (e.g., two atoms cannot occupy the same place, two covalent bonds cannot go through each other) and all the possible interactions (e.g., attractive or repulsive) between each component of the protein in each possible conformation, thus being able to calculate the free energy associated to each physically plausible conformation and eventually find the one with the lowest value. This approach spurred the expectation that, basically, the folding problem had already been solved in thermodynamic terms: in principle, finding the native state would just require exploring all the conformations (or enough of them) and calculating their free energy until the minimum is found. The protein folding problem was basically framed as a computational one.¹⁶ However, things turned out to be much more complex.

Consider a simplified protein with 100 amino acids (which is considered a relatively short length), in which every amino acid can only assume two different conformations. This protein has approximately 10^{30} potential conformations, with the native state corresponding, according to the thermodynamic hypothesis, to just one of them (or, more appropriately and realistically, to a very restricted set of these possible conformations). Moreover, if each conformational change occurred in just 1 picosecond, the folding process would take more time than that of the entire age of the universe if it were a totally random search (see Gomes & Fafsa, 2019 p. 26). However, proteins get folded in the order of milliseconds to seconds. What we are describing is a mental experiment known, in honour to its formulator, "Levinthal's

¹⁵Of course, assuming we possess an ideal model to tinker with (See Francoeur, 2001, 2002 for a historical revision of the use of models to study protein conformations).

¹⁶This computational approach is different from current computational models for predicting protein structures, such as AlphaFold (see note 6). In the case of the thermodynamic hypothesis, what is computed is the free energy of any possible 3D conformation that might be potentially acquired by a polypeptide chain based on knowledge concerning the physical basis of the interatomic interactions between its constituent amino acids. The only similarity between the two programmes is the neglect of kinetic considerations.

paradox” (Levinthal, 1968, 1969). Its morale is straightforward: the folding process cannot be a random process, otherwise it would take too long; the alternative is that there must exist some pathway guiding or biasing the conformational search from the unfolded to the native state, otherwise phenomena like the cooperative nature of the process (i.e., the seemingly all-or-none character of the transition between states) would remain unexplained.¹⁷ The most important implication of the thought experiment is that the folding process cannot be merely thermodynamically driven, as kinetic factors must play an essential role. In this respect, Levinthal’s postulation of folding pathways had the implication of reframing the protein folding problem by focusing on kinetic considerations. As Levinthal (1968, p. 44) argued: “... a pathway of folding means that there exists a well-defined sequence of events which follow one another so as to carry the protein from the unfolded random coil to a uniquely folded metastable state.” Another major consequence is that the native state does not necessarily correspond to a global energy minimum in the conformational energy landscape, but rather to the conformation that is most rapidly reachable (or slowest to exit from) from the unfolded protein. More stable conformations could exist, but as it is slower to get to them, it would be highly improbable for those to be reached. The native state might thus be characterised as a local minimum in the conformational energy landscape, a “metastable state”, i.e., a folded set of state(s) separated from the unfolded one by lower energetic barriers. Therefore, the mere search for energy minima would be rather irrelevant for explaining the folding process and to predict the native structure. The kinetic approach involved a change in the question guiding protein folding research (PFR): to explain the folding phenomenon and predict native structure, just computing the free energies of the possible conformations is insufficient as it is also necessary to describe the actual conformational changes (including the formation and breaking of the molecular interactions between different protein parts) that characterise each temporal stage during folding, starting from the unfolded state up to the native state. These theoretical developments gave rise to a new research agenda aimed to describe the folding pathways by characterizing the stages of the process, which were conceived as discrete and structurally characterizable intermediaries and transient states.¹⁸ This

¹⁷We return to the issue of cooperativity in Sect. 6.4.2.

¹⁸Let us clarify the concepts of intermediary, transient and transition state. An intermediary state refers to a discrete metastable conformation that can in principle be characterized thermodynamically. In kinetic terms, it refers to a conformation that lasts long enough to be “detectable directly”. Transient (or transitory) states refer to a mere stage in a dynamic process. A good analogy is with a car travelling from a city to another. An intermediary would be the car making a stop at a service station, and a transient state would be the car passing through any part of the road. The concept of transient state is somehow problematic because its status as a proper state is dubious and there could be cases in which the difference between being an intermediary or a transient state gets blurred. In biochemical practice, it is common for transient states to be conceptualised as discrete states representing the structural elements that are already formed at some stage of the folding process (like a picture of a car passing a specific point on the road). In what follows, we will assume this interpretation. A third concept is the transition state of a kinetic process, a theoretical concept of chemical kinetics referring to the transient state with highest energy in a reaction coordinate; this defines the so-called activation energy of a chemical reaction. In other words, in the

approach is currently labelled as the “classical view” of protein folding (Dill & Chan, 1997).

Anfinsen’s and Levinthal’s seminal contributions gave rise to two alternative sets of explanatory practices dealing with the protein folding phenomenon: a thermodynamic approach and a kinetic approach (or kinetic-dynamic approach). Both approaches are deeply intertwined in PFR but, as indicated above, there are important contrasts between them. Indeed, we would argue that these contrasts are profound enough to have divided the research field into different epistemic cultures. In the next section we shall analyse how both approaches account for the protein folding process.

6.3 Two Explanatory Approaches in Protein Folding Research

To understand the explanatory aims in PFR, in this section we describe how an “ideal explanation” of the protein folding process may look like for both the thermodynamic and kinetic approach. We characterize the concept of “ideal explanation” in the protein folding case as that accounting for an explanandum in terms of a complete description of the folding process given one specific set of epistemic resources concerning the physical and chemical properties and interactions at the atomic and molecular levels.¹⁹ This explanatory basis is largely common to thermodynamic and kinetic approaches (see Sects. 6.3.2 and 6.5.1). However, there are distinctive epistemic resources to each approach since, as we have already stressed, the former is centred on energetic and thermodynamic considerations, while the second focuses on kinetic considerations and structurally characterizable steps.

6.3.1 *Thermodynamic and Kinetic Explanations*

There are two main explananda in PFR: the first concerns the factors that determine the stability of the native structure; the second concerns how the protein acquires its native 3D structure starting from an unfolded state. We will call the first the native state stability problem (NSSP) and the second the folding dynamic problem (FDP).

protein case, this would be the least stable conformation a protein must transit through to reach native state (or to transit from any metastable state, like an intermediary, to another state), which defines the limiting step of the process. It is by definition non-detectable and non-isolatable, so it is only modelled theoretically. The aim of transition state theory was to account for the thermodynamic properties of chemical processes in terms of energy barriers between states.

¹⁹In biochemical practice, research results are often presented in terms of tokens. This is a deliberate idealization used for mainly narrative purposes. Consequently, the concept of ideal explanation is hereby characterised by reference to tokens; however, this is an analytical choice to account for the epistemic aims of the two explanatory approaches that will be addressed. Nevertheless, as we will show, biochemical practice is based on protein types.

In the case of the NSSP, an ideal thermodynamic explanation (i.e., leaving aside Levinthal's problem and assuming it is possible to identify all the possible conformations of a protein) would require the calculation of the free energy of all possible conformations of a polypeptide chain. The free energy of each one will be the result of the contribution of each (attractive and repulsive) interaction exerted between the protein parts in virtue of their intrinsic properties. The lower the conformational energy, the higher its stability. As anticipated in Sect. 6.2, the graphical representation of possible conformational energy states is called the energy landscape of the protein (Dill & Chan, 1997; Onuchic et al., 1997). Usually, it is illustrated as a graph in which the vertical axis represents the internal free energy and the other axes represent the conformational space.²⁰ Besides the possibility of finding the native conformation, knowing the shape of the energy landscape would also allow finding local energy minima, indicating the existence of possible alternative metastable conformations. Moreover, it would also be possible to describe energy barriers between the different metastable conformations, that is, the energies of the conformations that the protein should overcome to transition from one state to another. The higher the energy values of those intermediate conformations, the lower the probability of crossing from one state to another. In this way, a thermodynamic approach can account for the FDP. Importantly, in a thermodynamic explanatory approach there is neither an explicit appeal to the temporal variable nor to the actual pathways that a token protein (or populations of token proteins) will transit through when going from one state to another.

Conversely, the kinetic approach explicitly takes into consideration the time variable. The ideal explanation in this case is the description of the temporal development of a polypeptide chain when transiting from an unfolded state to the native state. This explanatory aim originates directly from Levinthal's postulation of folding pathways. In this case, what is explanatorily central are not the energy differences between conformational states, but the structural changes in conformation manifested by the developing protein during the folding process. This ideal explanation involves the characterisation of the temporal order in which the interactions between the parts of the protein occur and the identification of the new structures emerging as the native state is reached (Fersht, 1995, 1998; Baldwin, 2008; Englander & Mayne, 2017a, b). It is thus straightforward to see how the kinetic approach accounts for the FDP. Regarding the NSSP, from a kinetic perspective the stability of the native state is accounted for in terms of its maintenance. Basically, when the rate of reaching one state from another is higher in comparison to the rate of abandoning it, this state will be dynamically maintained.

²⁰More specifically, the vertical axis represents the 'internal free energy' and the additional axes represent the conformational space defined in terms of the conformational coordinates accounting for the degrees of freedom of a protein. The total internal free energy depends on physical factors (e.g., the sum of the energy contributions of hydrogen bonds, ion-pairs, torsion angles, hydrophobic contacts and salvation free energies) and the environmental factors on which the former depend (e.g., temperature, solvent). For each possible conformation, different interactions between protein parts will occur (or not), and that is what will define the total internal free energy of each conformation.

Although these two epistemic endeavours can be clearly distinguished conceptually, they are tightly intertwined in biochemical practice, so that it is usual to interpret kinetic features in thermodynamic terms and vice versa. The thermodynamic approach might explain kinetic features. For example, as we mentioned above, it is possible to describe the speed of folding/unfolding in terms of the height of the energetic barriers between the two states (using transition state theory): the higher the barrier, the slower the transition will be (Fersht, 1998). Conversely, as has already been related, the kinetic approach might explain thermodynamic features in terms of maintenance of states. For example, the stability of the native state can be explained by rapid refolding in contrast to a slow unfolding kinetic. This dynamic can be interpreted, for example, in terms of the early formation of strong stable interactions that “guide” the chain to the native state and which are then later difficult to break. In other cases, explanations of some specific features mesh thermodynamic and kinetic considerations, blurring their distinction. For example, the topology of the native state is sometimes assumed to play an important role both in the folding process and native state stability (Plaxco et al., 1998). A native fold with a complex topology (e.g., with very high contact order) will be reached more slowly than a native fold with a simpler one. But, from a thermodynamic point of view, a complex native state topology also increases the stability of a protein by the generation of atomic and molecular interactions constraining the unfolding process. Knotted topologies are an example of this latter case (Gomes & Faísca, 2019).

In summary, in various contexts both explanatory approaches are indeed intertwined and biochemists make a complementary use of their respective epistemic resources to generate explanations. Biochemists largely agree on the issue of which are the salient parts and activities in which the phenomenon under study should be decomposed to. However, as we shall argue in the next section, biochemists advocating thermodynamic and kinetic approaches engage in theoretical debates regarding the causal nature of their explanations, the explanatory relevance assigned to different aspects of the phenomena under study as well as on the status of some underlying ontological assumptions concerning the nature of the entities under study. These debates produce genuine clashes concerning both the interpretation of experimental results and the appropriate way to seek explanations of protein folding phenomena.

6.3.2 Mechanistic Credentials of Thermodynamic and Kinetic Explanations of Folding

To analyse to what extent both kinds of ideal explanations are causal, we will address the problem by considering to what extent they can be characterised as mechanistic explanations. The mechanistic framework of analysis is justified because of the ubiquitous appeal to underlying causes accounting for the

phenomena under study in PFR. To do this, a working definition of mechanism is needed. A largely consensual minimal definition of mechanism characterises the notion in terms of “entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon” (Glennan et al., 2022, p. 145). This implies that, in the case of protein folding, it would be necessary to identify the phenomenon to be explained, the parts and activities that are responsible for it and the organisation between the parts.

As we pointed out in Sect. 6.3.1, in the case of folding, the phenomena to be explained are at least two: the stability of the native state (NSSP) and the process of its acquisition (FDP). The second explanandum seems to be clearly amenable to mechanistic analysis. In fact, many standard characterizations of mechanism refer to an organised start-to-finish causal sequence of operations/activities performed by parts/entities producing a phenomenon (Machamer et al., 2000; Bechtel, 2011). However, mechanistic explanations do not only encompass start-to-finish causal sequences. Even when the maintenance of a state – another dynamical process, e.g., homeostasis – is at issue, a mechanistic explanation can be legitimately sought (Glennan et al., 2022). Given that kinetic explanations of both folding dynamics (which are classic examples of input-output aetiological explanations, see Krickel, Chap. 2, this volume) and maintenance of the native state are straightforwardly causal and mechanistic in nature, the causal nature of thermodynamic explanations for native state stability and acquisition will be a major concern in Sects. 6.5.1 and 6.5.2.

Concerning the relevant ontology of entities or parts, as indicated at the beginning of Sect. 6.3, there are aspects that are common to both explanatory approaches. Independently of whether the ideal explanatory aim is kinetic or thermodynamic, there is widespread agreement between the advocates of the kinetic and thermodynamics approach that the components of the polypeptide chain must be considered relevant parts in any explanation of native state stability and folding dynamics. These relevant parts must include the different covalently bonded atoms that compose the polypeptide chain, which are organized in the backbone as well as in the different residues of each amino acid. Accordingly, the relevant activities would then be accounted for in terms of the interactions established between these different parts, like hydrogen bonding, electrostatic interactions, VDW interactions, etc. which occur, at least partially (discounting relational extrinsic properties, see Santos et al., 2020) in virtue of the parts’ intrinsic properties, such as net electric charge, polar or non-polar nature, aromaticity etc. Other activities might be associated to the nature of the bonds established between its parts, like torsions between bonds, proline isomerization, steric clashes, chain collapse etc. Moreover, if we consider the environment in which the protein is embedded, other activities like the interaction with extrinsic factors like water, salts, protons, etc. may be considered. In many ways, all these parts and activities are common to both kinds of explanations.

6.4 Clashes Between Thermodynamic and Kinetic Approaches

Despite the general agreement just highlighted, there exists a clear difference between both approaches regarding the appropriate decomposition of folding phenomena. In this section we will analyse two main sources of disagreement. One pivotal source of this contrast concerns the explanatory relevance of the microscopic features of the folding process (Sect. 6.4.1). Another concerns the ontological commitments related to the decomposition of the system at hand (Sect. 6.4.2).

6.4.1 *Micro Versus Macro Analyses*

As we argued in Sect. 6.3.1, one first difference between thermodynamic and kinetic approaches concerns the appeal to the time variable. To put it bluntly, without countenancing the temporal aspect, it is difficult to see how thermodynamic explanations can be counted as causal. The rationale of the thermodynamic hypothesis is that reaching native conformation is dependent on exploring enough possible backbone conformations whose formation in turn depend on the intrinsic properties of the residues and peptide bonds. Obviously, this search takes time. However, from an ideal thermodynamic perspective, this temporal aspect would be explanatorily irrelevant to make sense of the directionality of the folding process and the stability of the native state. What is relevant is to account for free energy differences between the native state and the other physically possible conformations that the protein could attain. What is required is thus an explanation in terms of energetics. The issue of directionality in the thermodynamic approach is solved by assuming that the search for native state is (significantly) thermodynamically driven. In order to explain folding speed and cooperativity (Sects. 6.2 and 6.4.2), what is relevant are the energetic biases, which are represented by the currently proposed funnelled shapes of proteins' energy landscapes (Fig. 6.1). In addition to this, the peculiar idea that the energy landscape "directs the folding protein into the native state without the need for a definite pathway" (Govindarajan & Goldstein, 1998 p. 5545) or, put differently, that "native structure is determined only by the final native conditions" (Dill & Chan, 1997, p. 10), as if it were an attractor, are added. Basically, when folding starts, the number of possible conformations the protein can explore (i.e., the internal entropy of the chain) is gradually reduced due to the energetic biases accounted for by the enthalpic factors (such as the formation of intermolecular interactions) and the increment in solvent entropy as hydrophobic moieties get buried.²¹ In this respect, it is postulated that, considering a given protein type, folding may start at many different locations of the chain in each token's case, with the

²¹Even among the defenders of thermodynamic approaches there is considerable debate about which are the main factors that account for the energy biases of the folding process (Dill, 1999;

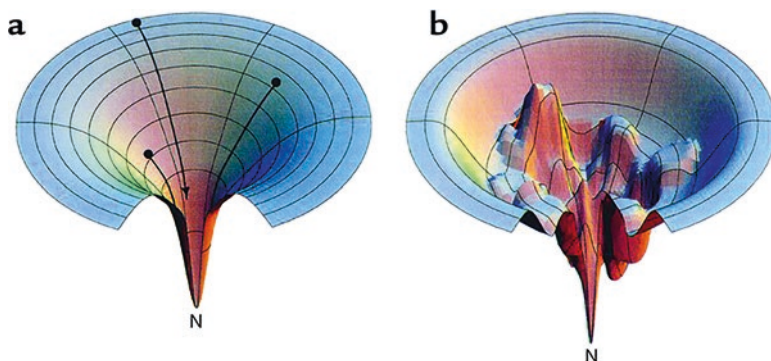


Fig. 6.1 Typical representations of energy landscapes of protein folding with funnelled shapes. The figure on the left corresponds to an idealized smooth funnel. Inside, three possible folding trajectories are marked as black lines starting from specific points located on the “denatured state ensemble”. The figure on the right corresponds to a rough and more realistic energy landscape with several local minima and energy barriers. (From Ken A. Dill, CC BY 4.0, via Wikimedia Commons)

further consequence that it will occur on many independent pathways (Fig. 6.1). Therefore, it becomes meaningless to postulate an order of folding events or the existence of single pathways to the native state.

This interpretation of folding dynamics is at odds with the notion of productive causal explanation because of the omission of the time variable and also because it focuses on energy differences instead of actual causal processes. This description thus leaves unexplained why, although all conformational potentialities can, in principle, be physically realised, some conformational potentialities are either not realised or transient (as suggested by experimental evidence); the thermodynamic explanation just assumes that it is because these conformations are energetically disadvantageous and unstable. Even at the microscopic level of description of the folding process (i.e., the level of the intrinsic properties of amino acids and peptide bonds, e.g., dihedral angles, side chain rotamers, etc.), which is the one explanatorily relevant for the thermodynamic account, the only concern is about differences in terms of stability between conformations, neglecting how the transition between conformations occurs. To solve this theoretical problem and make possible the explanation of aspects like the cooperativity of the folding process, the defenders of thermodynamic approaches resort to Brownian motion. Indeed, this dependence of folding on random processes makes folding analogous to a “parallel microscopic multi-pathway diffusion-like” process (Dill & Chan, 1997 p. 18), captured by the analogy between folding and the trickle of rainwater or skiers skying down a mountain, as is indicated in Fig. 6.1. This not only means that token proteins of the same type will inevitably fold differently, but that even the same token differently spatio-temporally localised (or even, at the extreme, the same spatio-temporally localised

Rose et al., 2006; Ben-Naim, 2015). Other driving forces in addition to the hydrophobic effect have been proposed, like the formation of hydrogen bonds, secondary structure propensities, etc.

token in case Brownian motion is an indeterministic process) will inevitably fold differently. This is why a central concept of the thermodynamic approach is the “denatured state”,²² which refers to an ensemble:

We can draw an analogy between the denatured ‘state’ and an ensemble of skiers distributed over a mountainside. When folding conditions are initiated, each skier proceeds down the funnel following his own private trajectory. Skiers skiing down funnels reach a global minimum (satisfying Anfinsen’s hypothesis) by many different routes (not a single microscopic pathway), yet they do so in a directed and rapid way (satisfying Levinthal’s concerns). Dill & Chan, 1997, p. 12.

From the thermodynamic perspective, the folding process is understood in terms of ensembles of different microscopic conformations. The concept of ensemble is difficult to characterize with precision due to its vagueness. In the case of protein folding, it can be conceptualised as the distribution of conformations that might be acquired by the token proteins of a population characterized by some macroscopic parameter (e.g., enthalpy differences, observable signals like fluorescence, etc.) and in a given environment (fixed temperature and pressure). This population is highly dynamic as each token protein is constantly fluctuating between different conformations: the broader the distribution of conformations, the higher the entropy and degrees of freedom of the population. The unfolded state would correspond to the ensemble with highest entropy. During folding, the entropy decreases until reaching native state, which corresponds to an ensemble with a very restricted conformational distribution. The distribution of conformations that define an ensemble is given by their stability differences, which, in their turn, are explained at the microscopic level by the interactions established between the protein components given their intrinsic properties.²³ The stages of protein folding are then conceived as ensembles with different conformational distributions in a protein population. Thus, assuming that the kinetic concept of pathway is only meaningful when referring to token and spatio-temporally localized polypeptide chains that actually start folding from one specific conformation, the thermodynamic approach denies the legitimacy of the kinetic (“classical view”) approach:

.... folding a protein does not involve starting from one specific conformation, A. The denatured state of a protein is not a single point on the landscape: it is all the points on the landscape, except for N. A pathway is too limited an idea to explain the flow from everywhere else, the denatured ensemble, to one point N. The concept of a pathway is useful for explaining the milestones we see in travels along a road or along a hiking trail, but not for describing how rain flows down a funnel. Dill & Chan, 1997 p. 12.

²²As stated in note 10, denatured is not the same as unfolded. In this case, the term “denatured” refers to any non-native state.

²³The energy landscape is thus said to “encode” the dynamic properties of the protein type. Encoding is due to the fact that the energy landscape represents the potency of a protein type, i.e., all possible conformations it may acquire given the intrinsic properties of the polypeptide chain’s components (in analogy with the fixed phase space of a dynamical system characterized in statistical mechanical terms). The energy landscape is therefore fixed from the outset (see Sect. 6.5.2).

Unlike the thermodynamic approach, kinetic approaches consider the time variable and aim to track the protein folding sequence of events through the identification of intermediate and transient states – structurally characterised – in the hope of uncovering the pathway leading to the native state. Folding dynamics are not random (otherwise, as Levinthal argued, the folding process would be too slow); they are rather constrained by processes (not necessarily thermodynamically driven)²⁴ that, despite being elusive, are open to experimental investigation. The discovery of such processes or principles of folding is the basic aim of kinetic approaches and what grounds their mechanistic ethos. Moreover, kinetic approaches do not deny that the folding pathways of different tokens of the same type of polypeptide chain might vary to some degree (if only because of Brownian motion). However, at some level of analysis, such pathways might share significant features, such as the generation of similar biochemically relevant intermediate and transient states, which can be described in structural terms by generalizing over the average behaviour of the same type of system (e.g., a population of tokens of the same protein type). A classic way to assess this is by treating the folding process as a chemical reaction going from the unfolded state (U) to the native state (N) and applying transition state theory to interpret kinetic experimental data. This permits modeling general features of the *transition state* of the process, which represents the highest energy state through which the protein must go through to transition from the unfolded to the native state. Using a previously determined structure of the native state (usually by X-ray crystallography) as a guide, and performing destabilizing site-directed mutations in the protein, it is possible to map the interactions which are already formed in this highest energy state, thus constructing a structural characterization of the limiting steps of the process (Fersht, 1995), which correspond to a global feature of a type representing the average behaviour of a protein population. This has allowed to propose different kinds of possible global mechanisms for protein folding, depending on which are considered the main events of the process (Fig. 6.2), e.g., formation and collision of secondary structure elements, hydrophobic collapse, nucleation-propagation, etc.²⁵

Advocates of the thermodynamic approach reject this macro-level kind of analysis as illegitimate for two reasons. First of all, at the atomistic level that is relevant for the thermodynamic approach, it is impossible that the pathways of two tokens can ever be identical (Eaton & Wolynes, 2017), if only because they are affected by thermal agitation. Secondly, the actual and extremely varied dynamics of folding tokens cannot be decomposed in terms of biochemically significant and structurally characterizable intermediate or transient states; the folding pathways for proteins of the same type uncovered by kinetic approaches will inevitably be too coarse-grained

²⁴This does not mean that they are in principle unexplainable in thermodynamic terms. What is important is that those processes are not merely driven (or accountable) by differences in stability and biased conformational search, but by sequences of causally related events.

²⁵Interestingly, an ongoing debate between the advocates of the kinetic approach concerns which kinds of mechanisms are most relevant and frequent to explain folding phenomena. See Gomes & Fafsa, 2019 p. 27–29 for an overview.

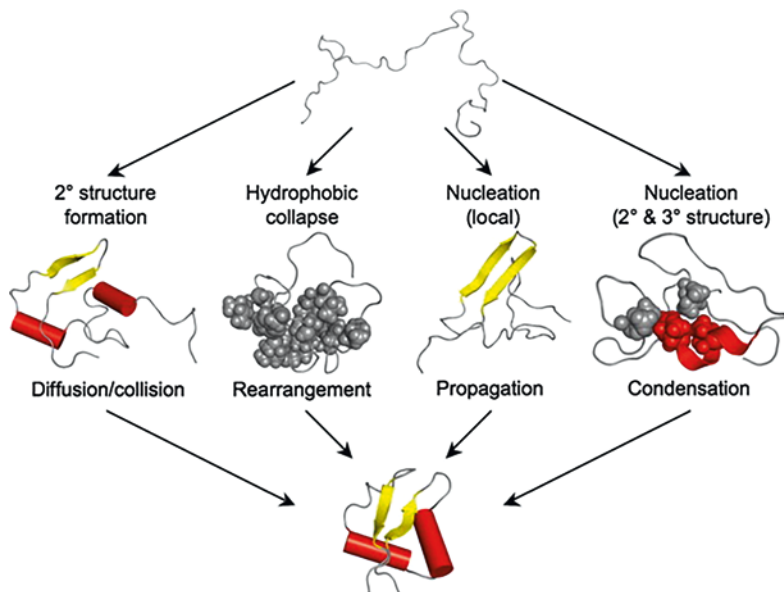


Fig. 6.2 Main kinds of folding routes that have been described for different proteins based on the structural characterization of the most relevant stages defining the folding process. (From Nickson and Clarke (2010), CC BY 3.0)

to ground significant generalizations. Ultimately, the thermodynamic approach denies the value of the structural characterization of the stages the protein transits through on the path to native state:

What is notable about the transition states of folding ... is not that they are specific structures, but that they are ensembles. The classical [kinetic] view focuses on specific structure (which experiments see), whereas the new view [thermodynamic]²⁶ is an ensemble perspective that recognizes the importance of disorder and that random processes and wrong steps are also major contributors to folding speed. Dill & Chan, 1997, p. 15.²⁷

This quotation, which is representative of the contrast characterizing current debates about folding, reveals a central issue. The characterization of the specific structures of intermediate and transient states makes theoretical sense in the context of models used to account for empirical data. In this context, for example, it is meaningful to treat the unfolded, intermediate and transient states as discrete populations. Meanwhile, random processes occurring at the microscopic level in ensembles of molecules can only be accounted for through theoretical representations such as ensembles and energy landscapes. This state of affairs illustrates the crucial point that one of the main clashes between thermodynamic and kinetic approaches

²⁶With “new view” these authors refer to the energy landscape theory.

²⁷The current form of the thermodynamic approach is commonly labeled as the “new view” by its advocates because of the novelty of the energy landscape theory and its application to protein folding. In this sense, it stands in opposition to the so-called “classical view” (see end of Sect. 6.2) based on the search for structurally characterizable stages in the folding process.

concerns the relative explanatory relevance that is attributed to empirical and theoretical considerations as well as to different theoretical models (pathway/sequential vs. landscape/parallel) and methods of analysis (structural, traditionally mechanistic, chemical kinetic vs. thermodynamic, statistical mechanical, chemical thermodynamics).

To summarise, the ideal thermodynamic explanation of folding dynamics aims to describe the whole space of possible conformations or potentials of any protein of a given type and the relative stability of each of them in terms of their free energy. The assumptions of the thermodynamic approach are that, while conformational search is dependent on micro-level causal factors, it is solely “constrained” thermodynamically (which also means that it is not totally random). In this sense, what is important in a thermodynamic explanation are microscopic level features, based on which protein’s type ensembles are defined. However, ensembles are not structurally, but thermodynamically characterized. Ultimately, the thermodynamic approach needs to explain in what specific causal sense thermodynamic “constraining” accounts for native state stability and folding dynamics. Section 6.5 shall delve on this issue, specifically on whether thermodynamic explanations of protein folding might be considered mechanistic or even causal. On the other hand, ideal kinetic explanations aim to uncover the temporal development of a polypeptide chain when transiting from the unfolded to the native state. The assumption of the kinetic approach is that this trajectory can be accounted for by describing the intermediate and transient states in structural terms. Ultimately, the kinetic approach needs to discover whether significant structural principles of folding exist notwithstanding variation in folding dynamics. The explanatorily relevant features are, then, macroscopic level properties corresponding to the average behaviours of polypeptide chains of the same type when transiting from the unfolded to the native state.

6.4.2 *The Issue of Decomposition*

Another general issue that emerges from the previous section is whether different kinds of analytic decompositions are possible. Despite the agreement between advocates of both approaches regarding the parts and activities composing the folding phenomena articulated in Sect. 6.3.2, the answer to this question is positive and shall be illustrated with one particular example: foldon kinetics (Englander & Mayne, 2017a, b).

Foldons might be defined as structural elements of a protein type acting as distinguishable cooperative units during the folding process (Fig. 6.3). Cooperativity means that the folding of one foldon influences the folding of the others, resulting in a stepwise folding process in which foldons acquire their native structure sequentially (i.e., when one foldon gets folded, this event triggers the folding of the next one and so on; conversely, a foldon cannot fold until the previous in the sequence gets folded first; more generally, a foldon is not stable enough on its own to last long enough unless the next foldon gets folded). Generally speaking, two elements of this different analytic decomposition are relevant. First, foldons – as relevant causal

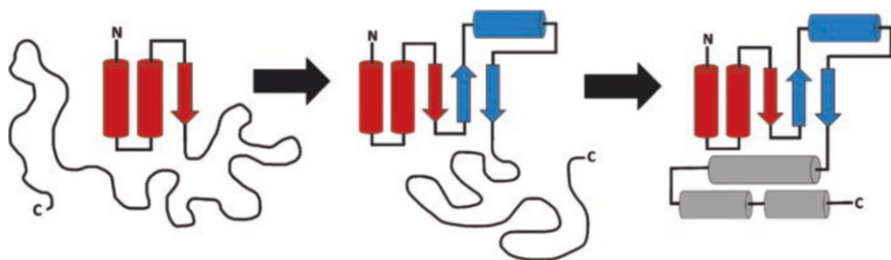


Fig. 6.3 Schematic representation of a sequential folding pathway containing three foldons (red, blue and grey). The scheme corresponds to a topological diagram in which α -helices are represented as cylinders and β -strands as arrows. Each stage corresponds to a transient state in which the folding of each foldon leads to the folding of the next one. The scheme represents the average behaviour of a protein type and not (necessarily) the actual pathway of a token protein. In this scheme the three foldons are concatenated in the protein sequence, but more complex cases (e.g., re-entrant topologies) are also possible. Additionally, this scheme takes foldons to be clusters of secondary structure elements, but other kinds of structural organizations (nucleations, hydrophobic centres, etc.) are not ruled out

parts with characteristic activities – cannot be identified at the initial stages of folding; they rather emerge as significant parts with specific causal roles during the transition from the unfolded to the native state. Thus, foldons provide a vivid example of the behaviour of dynamical systems with no fixed parts Levy and Bechtel (2016) refer to. Secondly, this example shows that the analytic decomposition into parts at the atomistic level is a commitment that is only strictly necessary for the thermodynamic approach. In fact, the denial on the part of the advocates of the thermodynamic approach that there are folding pathways is grounded on the assumption that conformational search occurs at the microscopic level: “*The multipathway idea stems from the early presumption that structure formation must occur through microscopic amino acid-level searching*” (Englander & Mayne, 2017b p. E9761). As we showed in the previous section, when looked at from this perspective, it becomes difficult to believe in significant folding pathways. Indeed, as Eaton and Wolynes (2017, p. E9759) admit: “*At an atomistic level, no two trajectories from the unfolded state to the folded state can possibly be identical, so there is an unimaginably large number of detailed pathways for folding a protein.*” The issue at this juncture is more significantly about the interpretation of the experimental evidence (gathered both *in vivo* and *in vitro*): the existence of pathway variation at the microscopic level is not under dispute, but its explanatory significance is at stake.

The foldon hypothesis stems from experimental approaches using new technologies (e.g., hydrogen exchange). Using these experimental approaches, foldon theorists have supposedly vindicated a series of kinetic hypotheses concerning the limited role (to the initial phases of folding) of random conformational search and, most prominently, the actual existence of biochemically relevant and structurally characterisable intermediates as well as the repeatable, stable, linear²⁸ and stepwise

²⁸This approach does not deny the possibility of parallel or ramified routes. Rather, the point is that these would consist, in their turn, of defined pathways with their respective structurally characterisable steps.

sequential nature of folding, at least in the case of some proteins (e.g., Rnase H): “... *Proteins fold by putting their structural elements into place over and over again in the same reproducible sequence*” (Englander & Mayne, 2017a, p. 8256). In particular, structurally characterizable transient states – i.e., “foldons” – pave the significant stepping stones of the folding process. This is the most important experimental finding in foldon research: foldons fold as units in sequential order (Englander & Mayne, 2017a, p. 8254). The foldon hypothesis suggests that conformational search is due to macro-level (i.e., not solely at the amino acid level) interactions between the components of foldons and between foldons, what Englander and Mayne (2017a) call cooperatively organized native-like intrafoldon and interfoldon interactions. In that sense, foldon research is centred on the structural characterization of the transient states that occur during the folding process, an endeavour that, to reiterate the point, is at odds with the thermodynamic approach to folding dynamics.

The foldon hypothesis gives rise to new research questions, such as: how do foldons and “foldon-based body plans” (Englander & Mayne, 2017a, p. 8256) evolve? What drives foldon assembly and interactions? For our present analytic purposes, the relevance of the foldon hypothesis is that, ultimately, the principles of folding are to be sought at the level of macro-entities such as foldons, cooperative units of amino acids, rather than atomistically. Probably the most significant ontological aspect of this view is that foldons in isolation are less stable than in the complex. This cooperativity between foldon components and between foldons suggests an anti-reductionist and relational view (see Santos, Chap. 12, this volume) of folding whereby the units of decomposition are macro-level structural or organizational units: a foldon cannot be characterised as just the sum of its components – i.e., amino acids – taken in isolation or, put differently, the behaviour of a foldon is not accountable in terms of the intrinsic properties of its components independently of their relational context. Thus, at least some version of the kinetic approach, such as the foldon hypothesis, should be contrasted to the thermodynamic approach in terms of their differing ontological commitments concerning the nature of the relevant entities and activities realizing the folding phenomena.

6.5 What Kind of Explanations Are Thermodynamic Explanations of Folding?

As we argued in Sect. 6.3.2, the kinetic approach aims to generate explanations of folding phenomena that are straightforwardly causal and mechanistic. Meanwhile, the causal nature of thermodynamic explanations remains dubious. In this section, we shall analyse how thermodynamic explanations of native state stability and folding dynamics may be interpreted.

6.5.1 *Thermodynamic Explanations of Native state Stability*

Our first attempt is to consider thermodynamic explanations of native state stability as instances of equilibrium explanations (Sober, 1983; Sperry-Taylor, 2019), in which the stable equilibrium condition that is maintained and to which the system returns when perturbed is, of course, the native state. According to Sober (1983), equilibrium explanations are not causal, as they do not refer to actual initial conditions or actual processes. However, Sober also argues that equilibrium explanations can be more informative than causal ones as they provide a particular kind of “understanding” related to situations whereby a system’s dynamical behaviour is governed by global equilibria, which are those to which a system reverts (or is “attracted to”) independently of initial conditions. In this case “...an event can be explained in the face of considerable ignorance of the actual forces and initial conditions that in fact caused the system to be in its equilibrium state. In this circumstance, we are, in one natural sense, ignorant of the event’s cause, but explanation is possible nonetheless” (Sober, 1983 p. 209).

However, this interpretation has some problems: when it is asked “why protein x reverts to putatively global equilibrium N ?”, are we not seeking a causal explanation? Consider this analogy with organismal homeostasis. For instance, internal temperature regulation is dependent on sensing external temperature; when the temperature increases, the organism responds (e.g., by producing some kind of metabolic change that, by assumption, is mechanistically accountable); thus, an explanation of heat regulation seems to be partially causal and mechanistic.

Note also that two explananda can be identified at this juncture: why a system reverts to equilibrium state and why the equilibrium state has that particular nature. What we are arguing here is that, in the case of homeostasis, we clearly rely on a causal and mechanistic explanans to explain equilibrium maintenance, even though in different organisms a different mechanism might act. The second explanandum (i.e., why internal temperature 37 °C is a global attractor or equilibrium) might have a different kind of explanation (e.g., evolutionary), which might nevertheless still be causal.

Analogous considerations, we surmise, pertain to native structure stability. In particular, just knowing the energy values of each possible conformation, which would give us the minimal energy value (or the maximally stable conformation, i.e., the native state), is neither enough to explain the dynamics of reversion to the native state nor to explain why the native state is the state of maximum stability. To achieve the first explanatory aim, it is necessary to appeal to the underlying mechanisms of stabilization in terms of the energy contributions of all the interactions formed between the parts of the protein, as well as the relations of the protein with the environment. Features like having a strong hydrophobic core, a great number of electrostatic interactions (e.g., salt bridges) on the surface, a high number of hydrogen bonds in certain configurations, etc. could explain why the native state of a protein is more stable than all (or most of) the other possible conformations. In the same way, the localization of a charged residue within the core, the existence of

hydrophobic patches at the surface, the existence of torsions, etc. could explain the instability of the native state of a given protein. In summary, to explain why the system reverts to the native state, it is necessary to appeal to the properties of the parts of the protein and the interaction between them and the environment. Dynamic reversion ($U \leftrightarrow N$) is grounded on structural properties (e.g., the existence and number of salt bridges, hydrogen and disulphide bonds), where proteins acquire such properties during folding in ways that are *prima facie* mechanistically accountable. Furthermore, given that the maintenance of the native state (including resistance to perturbation or return to equilibrium state after perturbation) is a dynamical process, kinetic considerations seem to be necessary (at least to complement ideal thermodynamic explanations) to account for proteins' behaviour.

To achieve the second explanatory aim – i.e., why the equilibrium state is that particular native state rather than another – an explanation might resort not only to principles of chemical stability but also to natural selection. Whether such principles are in principle mechanistically accountable is difficult to say.

Overall, the thermodynamic explanations of the two explananda related to native state stability (i.e., the reversion to equilibrium state and why to that particular equilibrium state) seem to us clearly amenable to be interpreted in causal and mechanistic terms, but only when complemented by either kinetic considerations (when the aim is to explain the maintenance of the native state) or chemical/evolutionary ones (when the aim is to explain the nature of the equilibrium state).

6.5.2 *Thermodynamic Explanations of Folding Dynamics*

When we consider thermodynamic explanations of folding dynamics, their mechanistic credentials are even more suspicious. First of all, thermodynamic explanations of folding dynamics seem intuitively more compatible with a formal deduction schema. They seem exemplars of the covering law model (by referring to a variety of thermodynamic generalisations and laws concerning Gibbs free energy, enthalpy and entropy) against which new mechanists have originally dedicated so much ink. However, as many authors have argued throughout history (see Santos, Chap. 12, this volume), even the purest mechanistic explanation must inevitably refer to some form of generalisation (a point more recently argued by Cartwright et al., 2020). We largely agree with this latter position and see no good reason to distinguish so sharply between mechanistic explanations and explanations referring to putative law-like generalisations. At the same time, advocates of the thermodynamic approach should clarify which thermodynamic generalizations are causal and, as such, proper explanantia for the phenomenon to be explained, that is, the acquisition of native structure.

Secondly, because of the absence of the temporal variable and the apparent omission of the causal details concerning how a protein reaches native state, it is difficult to make sense of the putative mechanistic (or even causal) nature of thermodynamic explanations. The lack of an explicit temporal and causal interpretation might

suggest that thermodynamic explanations are instances of constitutive mechanistic explanations. Indeed, since the energy landscape on which the ideal explanatory aims of this approach depends on is nothing more than the conformational possibility phase state of a protein type together with the constraints associated to the free energy of each conformation, it would be tempting to affirm that this description of the biochemical system constitutes (or is even identical to, see Craver et al., 2021) the phenomenon to be explained, that is, the thermodynamic behaviour of the protein. However, this constitutive interpretation would be at odds with the conception endorsed by practicing scientist advocating the thermodynamic approach, for whom the putative mechanisms underlying native state stability as well as those underlying folding dynamics cause rather than constitute the phenomena.

A third alternative interpretation is that explanations of folding dynamics are developmental explanations of a particular kind, that is, hybrids between constitutive and causal explanations. Following Ylikoski's (2013, p. 293) analysis, we might consider a developmental hybrid explanation of the following form: "Biochemical system S (i.e., the polypeptide chain type) has the causal capacity of folding to native state in a folding environment E due to S's components (i.e., the intrinsic properties of amino acids and peptide bonds) and their organization O (i.e., the arrangement of the amino acids in a linear polypeptide with a given sequence)." From a thermodynamic perspective, this explanation accounts for the fact that S generates different conformational distributions or ensembles at different stages of the process. However, thermodynamic approaches need to account for the directionality of the folding process in terms that are causally rich enough to make sense of its supposed "spontaneity", otherwise spontaneity is black boxed.²⁹ Whenever thermodynamic approaches provide causally rich details, for instance by appealing to the hydrophobic effect, they are implicitly committed to a causal account in terms of structural modifications, which is standardly mechanistic. This is indeed what kinetic approaches do. According to these, S as an unfolded polypeptide chain and S as a folded and functional protein differ in their organization, often in composition (e.g., if extrinsic components are integrated in the system) and, consequently, in their behaviour, as they acquire new properties as the folding process proceeds. The issue is thus not about constitution, but rather about whether thermodynamic approaches account for folding dynamics causally.

Another alternative to make sense of thermodynamic explanations of folding dynamics in causal terms should be mentioned. Often, the claims of the advocates of thermodynamic approaches seem to reify the geometrical properties of graphical representations such as energy landscapes, attributing causal roles to them that are difficult to comprehend. For instance, the claim according to which the conformational search is dependent on the "... *bias toward native interactions intrinsic to a funneled landscape*" (Eaton & Wolynes, 2017, p. E9759) gives rise to this kind of interpretation. Sometimes these claims are accompanied by others concerning the

²⁹The supposed spontaneity of the folding process has recently been the subject of criticism (see for example Sorokina et al., 2022).

directedness imparted to the search or its “encoding” (see note 23) by the energy landscape:

However, how this propensity [i.e., the thermodynamic bias] might be encoded in the physical chemistry of protein structure has never been discovered. One simply asserts the general proposition that it is encoded in the shape of the landscape and to an ad hoc principle named minimal frustration imposed by natural evolution. Englander and Mayne 2017a p. 8256

As this quotation illustrates, the reification of the properties of the energy landscape and the attribution of causal capacities to them is accompanied by assuming the biochemical relevance of debatable folding principles (e.g., minimal frustration)³⁰ as well as by a seemingly teleological interpretation of folding:

For effective performance, folding proteins must “know” how to select native as opposed to nonnative interactions. This information is said to be contained in the shape of the energy landscape, but how it is implemented in the physical chemistry of any given protein, or proteins in general, is unknown. Englander and Mayne, 2017a p. 8253

Therefore, the appeal to energy landscapes raises suspicions about the causal character of thermodynamic explanations of folding dynamics. What exactly do energy landscapes represent? A possible interpretation is that, since an energy landscape represents all the possible conformations a protein could attain, it also represents all the possible pathways or potential causal sequences a protein could take during folding. This interpretation would again lead us back to the issue concerning the causal nature of the processes underlying folding.

Unless we discount the causal legitimacy of thermodynamic explanations of folding dynamics *tout court*, there must be a way to bridge thermodynamic and causal talk. Indeed, in our opinion the most appropriate interpretation of thermodynamic explanations of folding dynamics is by bridging mechanistic analysis and energetics. This connects to the recent proposal that mechanistic explanations bottom out in energetics (Bechtel & Bollhagen, 2021). Basically, activities would be grounded on constraints on “the flow of free energy”. Thermodynamic explanations of folding emphasising energetic considerations approximate this new breed of extended mechanistic analysis focused on identifying the sources of free energy necessary for the mechanism to perform work and being active. In thermodynamic approaches of protein folding, Brownian motion must be countenanced as a significant force. In a sense, it might be considered as one source of activity or free energy underlying folding. However, if Brownian motion solely regulated the folding process, we would be left wondering how native state can as a matter of fact be reached. Therefore, to answer the question of why some atomistic interactions tend to occur

³⁰Frustration refers, in this context, to a general property of any linear chain composed of monomers of different nature like, for example, a random peptide chain. In the case of a random peptide chain, the most common fate will be a collapse in many different 3D structures instead of a unique (or very restricted ensemble of) globular, stable and soluble 3D structures. Nevertheless, extant proteins do most often fold into a native state. This behaviour is called minimal frustration: the fact that extant proteins acquire native state would be the result of the selection of “minimally frustrated” peptide chains through evolution. This “minimal frustration principle” is at the core of current versions of the thermodynamic hypotheses.

with higher probability than others, what is needed is a more rigorous causal account of the conformational search in thermodynamic terms. One possible suggestion is that the differences in stability (i.e., accounted in terms of Gibbs free energy) between the different conformations determine which ones are transient and which ones are acquired during the process leading to native state. However, as Englander and Mayne (2017a, p. 8257) note, this way of interpreting the thermodynamic constraint does neither refer to nor identify any relevant molecular properties of proteins:

Atypically, the funneled landscape emblematic of energy landscape theory does not deal with molecular properties that would serve to guide interactions. It portrays some external thermodynamic constraints that are valid for the folding of proteins, RNA, or any other polymer. It contains in itself no molecular information or molecule-based constraints or predictions.

If the energetic considerations considered central to folding dynamics by advocates of the thermodynamic approach are so general as to pertain to any polymer, it is not surprising that a causal and mechanistic interpretation of the thermodynamics of folding dynamics is not easily forthcoming.

In a sense, this difficulty is not surprising, as thermodynamic approaches are grounded on disciplines such as chemical thermodynamics (whose primary focus is on the direction of chemical reactions independently of the underlying reaction mechanisms) and inspired by statistical mechanics. One source of inspiration is the analogy with ideal gases, that captures the important point that, despite continuous change at the micro-level (due to Brownian motion for instance), the ensemble during folding changes only in terms of the probability distribution of the fixed set of possible protein conformations – which are analogous to microstates (see Sect. 6.4.1 and note 23). This analogy is questionable in many senses. First, the uniformity assumption (i.e., with the idea that the molecules of an ideal gas are identical) is questionable in the protein case; in fact, proteins of a same type might vary in composition, for instance by acquiring new structural components from the environment or even by acquiring variations in composition that might occur in-vivo because of mistranslation. Secondly, as advocates of kinetic approaches stress, there is no reason to believe, experimentally and theoretically, that all possible conformations of a protein will as a matter of fact occur during folding. In the biochemical case, some conformations might never be realized. Thirdly, as advocates of kinetic approaches also stress, there is no experimental reason to deny that only some conformations are biochemically significant. To argue for the contrary position, as advocates of the thermodynamic approach do, is to borrow uncritically the analogy with ideal gases, where all microstates are assumed to be (“in the long term”) equiprobable (i.e., ergodic hypothesis of statistical thermodynamics). Fourthly, and most importantly, unlike phase space, the energy landscape might not be fixed (Sect. 6.4.1) at the outset by the intrinsic properties of the protein type, grounded on its characteristic amino acid composition. It is rather co-determined by the properties of the protein environment, including the varieties of molecules the developing protein interacts with during folding (see Sorokina et al., 2022). In a nutshell, unlike

phase space, the energy landscape might be dynamic and, consequently, cannot be reduced to a mere representation of the polypeptide chain's degrees of freedom considered as an isolated system (i.e., solely characterized in terms of the intrinsic properties of its components).³¹

At the same time, as already indicated in Sect. 6.5.1, thermodynamic approaches make implicit reference to causal processes such as the formation of salt bridges, hydrogen bonds or hydrophobic effects. In this sense, note that the hydrophobic effect can either be given an energetic interpretation (whereby, in an aqueous medium, the solvent entropy is higher when hydrophobic moieties are not being solvated, i.e., the free energy of the system is lower, so that they are grouped together) or a mechanistic interpretation (whereby hydrophobic amino acids tend to move inside the folding structure while hydrophilic ones tend to position themselves on the external part of the protein structure during the folding process, leading to the compaction of the polypeptide chain). Both interpretations have a sound rationale and might be considered, as it is often the case in biochemical practice, complementary. The contrast between thermodynamic and kinetic approaches might as a result be more properly diagnosed as a dispute concerning the appropriate analytic strategy to explain folding phenomena.

6.6 Conclusion

The protein folding problem is a philosophically fertile field that has received, as far as we know, limited attention in the philosophies of chemistry and biology despite its central importance in biochemistry and, more generally, biology. There are many aspects of this problem that remain to be addressed beyond those considered in this chapter. In this sense, it should be stressed that part of the contrast between thermodynamic and kinetic approaches is intimately related to the use of different experimental and modelling techniques. Advocates of the kinetic approach tend to base their arguments mostly on experimental results, while advocates of the thermodynamic approach tend to adopt mostly theoretical and computational practices, being computer simulations of simplified models a central one. At the same time, both approaches have common difficulties when attempting to explain why extant proteins fold as they do, independently of whether the answer is sought by asking the question of why a protein has a particular energy landscape or, alternatively, why a specific order of stages occurs. When faced with these deep questions, both approaches often resort to evolutionary biology. The most common answer, which

³¹Analysing the case of allostery, Neal (2021, p. 209) indeed argues that the analogy with ideal gases is broken because "The perturbation of the ensemble by the allosteric ligand remodels the energy landscape of the entire system because the energetic properties of the microstates themselves were differentially altered by the allosteric ligand." It remains unclear whether Neal is arguing that this is just a change in the probability distribution of microstates or whether a new set of microstates is formed in the process.

betrays an adaptationist bias (see the principle of minimal frustration, note 24), is that extant proteins are as they are and fold as they fold because they are the results of adaptive evolution. In other words, extant proteins are assumed to be adaptive traits that can reach certain conformations in their physiological contexts at a speed that allows them to perform their biological functions.

It must also be noted that, despite its relative antiquity, the protein folding problem remains an unsolved problem in biochemical and biophysical research. Despite experimental and theoretical advances (including data-driven approaches such as AlphaFold, see note 6), new questions and new debates are continuously emerging. In this article we have assessed one major source of disagreement, rooted in the divergence between two major explanatory approaches to the protein folding problem that can be traced back to when the problem was firstly formulated: a thermodynamic approach focused on energetic features and a kinetic approach centred on temporal and structural ones. Despite the partial agreement between these approaches on various aspects of the folding process and their complementarity – evident in biochemical practice – in generating hybrid explanations, there remain significant contrasts between them. We have tried to uncover some aspects of such contrast related to the relevance assigned to different epistemic resources and the causal nature of the explanations proposed. We also identified their different ontological assumptions. While the thermodynamic approach localizes the relevant epistemic resources at the atomic level, thus aiming to define the properties of ensembles, the kinetic approach considers as central the average behaviour of populations of proteins, thus aiming to provide a structural description of the stages of the folding process. Thermodynamic approaches are based on a conceptualization of the folding process whereby pathways are unwarranted postulations. Conversely, kinetic approaches deny the relevance of the atomic level of analysis (by itself) because it is experimentally inaccessible. Concerning the causal nature of the explanations, kinetic explanations are straightforwardly causal and mechanistic, while the causal nature of thermodynamic ones is elusive and difficult to interpret. Here the incipient contrast between thermodynamic and structural approaches comes to the fore, which is a further expression of the difficulty of applying a structure-based mechanistic model of explanation to chemistry (Scerri, Chap. 8, this volume) and physics (Falkenburg, Chap. 10, this volume) alike.³² Finally, the thermodynamic approach conceives the folding process as a manifestation (or suppression) of the predetermined intrinsic properties of the protein components, while the kinetic approach – at least in some forms – seems compatible with an anti-reductionist and relational perspective (Santos, Chap. 12, this volume) whereby proteins, during development, diachronically acquire novel properties, some of which only characterizable macroscopically. What results from this state of affairs is a “hardly-to-integrate” pluralism (Bolinska, 2022 reaches similar conclusions in the case of protein structure determination). At first glance, these disagreements may be interpreted as a classical

³² Structural approaches in biochemistry become especially problematic when protein function is not dependent on structural change, as IDPs (see note 7) seem to show.

example of “relative significance” debate (Beatty, 1997). However, in this case both approaches have led to interpretations of natural phenomena that are difficult to reconcile. Indeed, from the thermodynamic perspective, it is always possible to argue that structural pathways are chimeras produced by centring attention on selected macroscopic features. Conversely, for the advocates of the kinetic approach, it is always possible to argue that multiple and parallel folding routes are irrelevant as a basis for generalizable explanations. These contrasts are in our opinion hardly resolvable.

Acknowledgements Gabriel Vallejos-Baccelliere acknowledges the financial support of Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT grant N° 3210758). Davide Vecchi acknowledges the financial support of the FCT—Fundação para a Ciência e a Tecnologia (DL57/2016/CP1479/CT0072; Grants N. UIDB/00678/2020 and UIDP/00678/2020). We would like to thank Gil Santos, Eric Scerri and two anonymous reviewers for stimulating criticisms and suggestions to improve the clarity of the manuscript. Additional thanks to Ariel Roffé, Santiago Ginnobili, Daniel Nicholson, Luis González-Flecha and the audience at the 17th CLMPST, held in Buenos Aires, Argentina. Gabriel Vallejos-Baccelliere also thanks the academic staff of Laboratorio de Bioquímica y Biología Molecular, Departamento de Biología, Facultad de Ciencias, Universidad de Chile.

References

- Austin, R. C. (2009). The unfolded protein response in health and disease. *Antioxidants & Redox Signaling*, *11*(9), 2279–2287. <https://doi.org/10.1089/ars.2009.2686>
- Baldwin, R. L. (2008). The search for folding intermediates and the mechanism of protein folding. *Annual Review of Biophysics*, *37*, 1–21. <https://doi.org/10.1146/annurev.biophys.37.032807.125948>
- Beatty, J. (1997). Why do biologists argue like they do? *Philosophy of Science*, *64*, S432–S443. <http://www.jstor.org/stable/188423>
- Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, *78*(4), 533–557.
- Bechtel, W., & Bollhagen, A. (2021). Active biological mechanisms: Transforming energy into motion in molecular motors. *Synthese*, *199*(5–6), 12705–12729.
- Ben-Naim, A. (2015). *Myths and verities in protein folding theories*. WSPC.
- Bolinska, A. (2022). Monist proposal: Against integrative pluralism about protein structure. *Erkenn*. <https://doi.org/10.1007/s10670-022-00601-2>
- Cartwright, N., Pemberton, J., & Wieten, S. (2020). Mechanisms, laws and explanation. *European Journal of Philosophy of Science*, *10*, 25. <https://doi.org/10.1007/s13194-020-00284-y>
- Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, *199*(3–4), 8807–8828. <https://doi.org/10.1007/s11229-021-03183-8>
- Dill, K. A. (1999). Polymer principles and protein folding. *Protein Science: A Publication of the Protein Society*, *8*(6), 1166–1180. <https://doi.org/10.1110/ps.8.6.1166>
- Dill, K. A., & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Structural Biology*, *4*(1), 10–19.
- Eaton, W. A., & Wolynes, P. G. (2017). Theory, simulations, and experiments show that proteins fold by multiple pathways. *PNAS*, *114*(46), E9759–E9760.
- Englander, S. W., & Mayne, L. (2017a). The case for defined protein folding pathways. *PNAS*, *114*(31), 8253–8258.
- Englander, S. W., & Mayne, L. (2017b). Reply to Eaton and Wolynes: How do proteins fold? *PNAS*, *114*(46), E9761–E9762.

- Fersht, A. R. (1995). Characterizing transition states in protein folding: An essential step in the puzzle. *Current Opinion in Structural Biology*, 5, 79–84.
- Fersht, A. R. (1998). *Structure and mechanism in protein science: A guide to enzyme catalysis and protein folding*. W. H. Freeman.
- Francoeur, E. (2001). Molecular models and the articulation of structural constraints in chemistry. In U. Klein (Ed.), *Tools and modes of representation in the laboratory sciences* (Boston studies in the philosophy and history of science) (Vol. 222). Springer. https://doi.org/10.1007/978-94-015-9737-1_7
- Francoeur, E. (2002). Cyrus Levinthal, the Kluge and the origins of interactive molecular graphics. *Endeavour*, 26(4), 127–131. [https://doi.org/10.1016/s0160-9327\(02\)01468-0](https://doi.org/10.1016/s0160-9327(02)01468-0)
- Glennan, S., Illari, P., & Weber, E. (2022). Six theses on mechanisms and mechanistic science. *Journal for General Philosophy of Science*, 53, 143–161.
- Gomes, C., & Faisca, P. (2019). *Protein folding: An introduction*. Springer.
- Govindarajan, S., & Goldstein, R. A. (1998). On the thermodynamic hypothesis of protein folding. *PNAS*, 95(10), 5545–5549.
- Krebs, J. E., Goldstein, E. S., & Kilpatrick, S. T. (2018). *Lewin's genes XII*. Jones & Bartlett Learning.
- Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65, 44–45.
- Levinthal, C. (1969). How to fold graciously. Mossbauer spectroscopy in biological systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois: 22–24.
- Levy, A., & Bechtel, W. (2016). *Towards mechanism 2.0: Expanding the scope of mechanistic explanation*. <http://philsci-archiv.pitt.edu/12567/>
- Liu, P. P., Xie, Y., Meng, X. Y., & Kang, J. S. (2019). History and progress of hypotheses and clinical trials for Alzheimer's disease. *Signal Transduction and Targeted Therapy*, 4, 29. <https://doi.org/10.1038/s41392-019-0063-8>. Erratum in: *Signal Transduct Target Ther*. 2019 Sep 23;4:37. PMID: 31637009; PMCID: PMC6799833.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67(1), 1–25.
- Medina, E., R Latham, D., & Sanabria, H. (2021). Unraveling protein's structural dynamics: From configurational dynamics to ensemble switching guides functional mesoscale assemblies. *Current Opinion in Structural Biology*, 66, 129–138. <https://doi.org/10.1016/j.sbi.2020.10.016>
- Neal, J. P. (2021). *Protein structure, dynamics, and function: A philosophical account of representation and explanation in structural biology*. Ph.D. Dissertation, The University of Pittsburgh, USA. <http://d-scholarship.pitt.edu/41715/13/NealJP%20ETD%20History%20%26%20Philosophy%20of%20Proteins.pdf>
- Nickson, A. A., & Clarke, J. (2010). What lessons can be learned from studying the folding of homologous proteins? *Methods (San Diego, Calif.)*, 52(1), 38–50. <https://doi.org/10.1016/j.ymeth.2010.06.003>
- Onuchic, J. N., Luthey-Schulten, Z., & Wolynes, P. G. (1997). Theory of protein folding: The energy landscape perspective. *Annual Review of Physical Chemistry*, 48, 545–600. <https://doi.org/10.1146/annurev.physchem.48.1.545>
- Plaxco, K. W., Simons, K. T., & Baker, D. (1998). Contact order, transition state placement and the refolding rates of single domain proteins. *Journal of Molecular Biology*, 277(4), 985–994. <https://doi.org/10.1006/jmbi.1998.1645>
- Rose, G. D., Fleming, P. J., Banavar, J. R., & Maritan, A. (2006). A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 103(45), 16623–16633. <https://doi.org/10.1073/pnas.0606843103>
- Santos, G., Vallejos, G., & Vecchi, D. (2020). A relational-constructionist account of protein macrostructure and function. *Foundations of Chemistry*, 22(3), 363–382.
- Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43(2), 201–210.

- Sorokina, I., Mushegian, A. R., & Koonin, E. V. (2022). Is protein folding a thermodynamically unfavorable, active, energy-dependent process? *International Journal of Molecular Sciences*, 23, 521. <https://doi.org/10.3390/ijms23010521>
- Sperry-Taylor, A. T. (2019). Reassessing equilibrium explanations: When are they causal explanations? *Synthese*, 198(6), 5577–5598.
- Tanford, C., & Reynolds, J. A. (2003). *Nature's robots: A history of proteins*. Oxford University Press.
- Vallejos-Baccelliere, G. (2022). Problemas contemporáneos en la filosofía de la bioquímica. *Culturas Científicas*, 3(1), 45–77. <https://doi.org/10.35588/cc.v3i1.5584>
- Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis*, 78(2), 277–297.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 7

Mechanisms in Chemistry



Robin Findlay Hendry

Abstract Mechanisms are the how of chemical reactions. Substances are individuated by their structures at the molecular scale, so a chemical reaction is just the transformation of reagent structures into product structures. Explaining a chemical reaction must therefore involve different hypotheses about how this might happen: proposing, investigating and sometimes eliminating different possible pathways from reagents to products. One distinctive aspect of mechanisms in chemistry is that they are broken down into a few basic kinds of step involving the breaking and making of bonds between atoms. This is necessary for chemical kinetics, the study of how fast reactions happen, and what affects it. It draws on G.N. Lewis' identification of the chemical bond as involving shared electrons, which from the 1920s achieved the commensuration of chemistry and physics. The breaking or making of a bond just is the transfer of electrons, so a chemical bond on one side of an equation might be balanced on the other side by the appearance of a corresponding quantity of excess charge. A bond is understood to have been exchanged for a pair of electrons. Since reaction mechanisms rely on identities, doesn't the establishment of a reaction mechanism explain away the chemical phenomena, showing that they are no more than the movement of charges and masses? In one sense yes: these mechanisms seem to involve a conserved-quantity conception of causation. But in another sense no: the 'lower-level' entities can do what they do only when embedded in higher-level organisation or structure. There need be no threat of reduction.

Keywords Chemistry · Classification · Emergence · Mechanisms · Reduction

7.1 Introduction

In this paper, I will defend the following claims which, taken together, amount to a detailed conception of the metaphysics of mechanisms in chemistry.

R. F. Hendry (✉)

Department of Philosophy, Durham University, Durham, UK

e-mail: r.f.hendry@durham.ac.uk

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_7

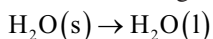
139

1. Chemical reactions are processes individuated by the molecular-scale structures with which they begin (the reagents) and end (the products).
2. A reaction mechanism is a description of *how* a chemical reaction might possibly happen: *how* the reagents are transformed into the products.
3. Different pathways from reagents to products are individuated by how they are composed of a few basic kinds of step involving transfers of such conserved quantities as mass, charge and energy.
4. This view of mechanisms has no tendency to support the reduction of chemistry to physics, or reductionism more generally.

My intention in this article is not to say anything complicated or controversial about reaction mechanisms in chemistry, but instead to relate some obvious and basic facts about them to the recent wave of philosophical literature on mechanisms. Even though chemistry has a long and illustrious tradition of thinking about mechanisms – in the twentieth century, multiple Nobel prizes were awarded for work on mechanisms – mechanist philosophers have almost entirely looked elsewhere for examples. Thus, for instance, the word ‘chemistry’ does not occur in the influential paper that initiated that wave (Machamer et al., 2000). In the major survey article by Craver and Tabery (2016), chemistry is mentioned only once on its own account (that is, as a science that investigates mechanisms), and then only as a subject to be addressed in future work. I think there are some distinctive and interesting things to say about reaction mechanisms in chemistry.

7.2 What Is a Chemical Reaction?

A chemical reaction is a type of process in which chemical change occurs.¹ Chemical reactions can be specified at two different scales, or levels:² as involving chemical substances, or as involving species at the molecular scale (atoms, ions or groups of atoms). Chemical change is not just any change: for a chemical reaction to have taken place during a process, the substances or species at the end must be chemically different from those at the beginning. The melting of ice does not count as a chemical reaction, although it does, of course, have a mechanism:



Given all the above, it seems obvious that chemical reactions are individuated by the substances or species with which they begin (the reagents) and those with which they end (the products). Chemical reactions can be specified at different levels of abstraction. Consider for instance the reaction between hydrogen chloride and

¹Everything I say in this section is intended to be consistent with relevant definitions agreed by the International Union of Pure and Applied Chemistry (IUPAC: see <https://goldbook.iupac.org>)

²I have no major objection to using levels, but in this paper I will opt for scales because (i) scales will do the job, and (ii) they are free of some distracting philosophical connotations associated with levels, which some metaphysicians fear give rise to important confusions in the context of debates about reduction and emergence.

sodium hydroxide to form sodium chloride and water. From there one might abstract to yield a reaction type in which an unspecified hydrogen halide reacts with an unspecified metal hydroxide to give a metal halide and water. Yet more abstractly, one might consider as a class reactions between acids and bases to give a salt and water. The same is true in organic chemistry: one might consider the oxidation of propan-2-ol ($\text{CH}_3\text{CHOHCH}_3$) to propanone (CH_3COCH_3 , better known as acetone), or more abstractly the oxidation of secondary alcohols of the general form R_1CHOHR_2 to ketones of the general form R_1COR_2 , with R_1 and R_2 being unspecified alkyl groups. A description of a chemical reaction always picks out a reaction type, but the level of abstraction of that description may vary.

Chemists give information about chemical reactions using representations of the relevant processes, which are often, but not always, balanced chemical equations. These involve chemical formulae, which are themselves representations of the structure of the relevant chemical substances, which again can be couched at different levels of detail (or abstraction). The simplest kind of chemical formula – the empirical formula – represents only the elemental composition of a substance. Because the information it provides about structure³ is limited, this formula will fail to distinguish between distinct substances. For instance, acetone, mentioned earlier, has the empirical formula $\text{C}_3\text{H}_6\text{O}$, which is shared by about 10 distinct substances including cyclopropanol and ethanal (the substance formerly known as acetaldehyde). Empirical formulae are not very informative: for any purpose that requires isomers be distinguished, more detailed structural formulae will need to be used.

I have argued elsewhere for microstructural essentialism, the thesis that chemical substances are the substances they are in virtue of their structures at the molecular scale (see Hendry, 2023, [Forthcoming](#), from which the following discussion is drawn). I will not argue here for full-blown microstructural essentialism but only the weaker claim that microstructuralism is the official ideology of chemistry, suffusing its approach to mechanisms. I will do that via three arguments drawing on the practice of chemistry, concerning (i) the centrality of structure at the molecular scale to chemical classification and nomenclature, and the complete absence of any other *non*-microstructural criteria; (ii) the role of microstructure in explaining and predicting the chemical and physical behaviour of substances, and (iii) the fact that no other systematic basis for individuating substances is consistent with chemical practice, and the epistemic interests that underlie it. I will develop those three arguments in turn.

The case for microstructuralism concerning chemical classification and nomenclature is particularly strong in the case of the chemical elements. Since 1923, the International Union of Pure and Applied Chemistry (IUPAC) has quite explicitly identified nuclear charge as what characterises the various chemical elements (for the historical background see van der Vet, 1979; Kragh, 2000). I have argued that the historical record supports realism about the elements as natural kinds, because the IUPAC change reflected a series of discoveries (Hendry, 2006, 2010a). I

³Here I am using ‘structure’ inclusively, to include elemental composition. Structure is sometimes contrasted with composition, but to know that a particular substance is composed of certain elements in certain proportions is to know something about its structure at the molecular scale.

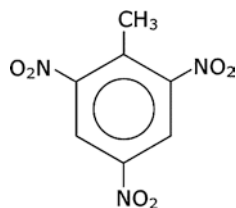
correspondingly disagree with LaPorte (2004, Chapter 4), who has argued that, prior to the twentieth century, it was indeterminate whether the names of the chemical elements referred to classes of atoms which are alike in respect of their nuclear charge, or to classes of atoms alike in respect of their atomic weight, or to classes of atoms alike in both respects. For LaPorte, IUPAC's 1923 decision had the character of a stipulation. I think it is quite natural to see the extensions of the names of the elements as being determinate before 1923, and IUPAC's identification as simply the recognition of determinate membership (see Hendry, 2006, 2010a). Silver, which has been known since ancient times, always consisted of roughly equal mixtures of two isotopes of silver (^{107}Ag and ^{109}Ag), differing in respect of their atomic weight. There is, of course, a remote possibility that the isotopic composition of silver changed radically over time, but that would not affect my main point, which is that the weight differences between the isotopes make very little difference to their chemical behaviour. What makes those diverse atoms count as silver is what they share, namely their nuclear charge (47), a property that explains why these diverse atoms behave chemically in very similar ways. The twentieth-century identification of nuclear charge as what individuates the elements was a discovery of this fact, rather than a stipulation or a convention.

There are good grounds for extending microstructuralism to *compound* substances. The rules for chemical nomenclature that IUPAC has developed over the years are based entirely on microstructural properties and relations. Consider for instance 2, 4, 6,-trinitromethylbenzene, better known as trinitrotoluene, or TNT (see Fig. 7.1).

For the purposes of nomenclature this compound, which was first synthesised in the nineteenth century, is regarded as being derived from methylbenzene (toluene): counting clockwise from the methyl ($-\text{CH}_3$) group as position 1, there are three nitro-groups ($-\text{NO}_2$) at positions 2, 4 and 6, replacing three hydrogen atoms (conventionally the remaining hydrogen atoms at positions 3 and 5 are left out for clarity). In short, TNT is named purely on the basis of its bond structure. Now it is true that there are alternative ways of generating names for compounds (see Leigh et al., 1998), but the important point is that IUPAC's various systems of nomenclature are all based on microstructure. Critiques of microstructuralism in chemistry never seem to mention this fact.

A second argument for microstructuralism concerns explanation. Understanding the chemical behaviour of a compound substance—its chemical reactivity—is essentially a matter of understanding how its structure transforms into the structure of other substances under various conditions. As we shall see, this involves the

Fig. 7.1 Full structural formula for 2, 4, 6,-trinitrotoluene, or TNT



study of chemical reaction mechanisms. Similarly, understanding the physical behaviour of a substance, including its melting and boiling points and its spectroscopic behaviour, is a matter of understanding how, given their structure, its constituent molecules interact with each other and with radiation respectively.

A third argument for microstructuralism concerns the fact that no other group of properties provides a systematic framework for naming and classifying substances, or understanding their behaviour, that is consistent with chemical practice. What are the alternatives? Paul Needham (2011) has argued that classical thermodynamics provides macroscopic relations of sameness and difference between substances, acknowledging that his macroscopic perspective is revisionary of current chemical practice, in that it divides substances more finely than chemists do. On Needham's view, different isotopes of the same element are distinct substances even though chemists lump them together when thinking about the elements. Taking a similar stance to LaPorte, although for different reasons, Needham describes IUPAC's identification of nuclear charge as what characterises the elements as a 'convention' (2008, 66). I think this is too thin a description, and historically misleading. The adoption of nuclear charge was a considered choice, which reflected the *discovery* that the elements occur in nature as mixtures of different isotopes. This meant that the names of the elements *as they were currently being used* had been *discovered* to refer to populations of atoms which are alike in respect of their nuclear charge, while diverse in respect of their weight. The identification of nuclear charge as what characterises the elements was therefore simply a recognition of the real basis of the periodic table (see Hendry, 2006, 2010a). Further objections to Needham's thermodynamic criteria are provided by other examples. Orthohydrogen and parahydrogen are spin isomers of the hydrogen molecule that readily interconvert on thermal interaction: in orthohydrogen the spins are aligned, while in parahydrogen they are opposed. Orthohydrogen and parahydrogen count as different substances on the thermodynamic criteria. The same holds for populations of atoms in mutually orthogonal quantum states, such as two streams of silver atoms emerging from a Stern-Gerlach apparatus (for detailed discussion see Hendry, 2010b). A natural conclusion is that Needham's proposed thermodynamic criteria for sameness and difference of substance track differences of physical state, without regard for whether those differences correspond to distinctions of substance.⁴

These three arguments together constitute a strong positive case in favour of taking microstructuralism to be the prevailing classificatory ideology of the discipline of chemistry, and seeing chemistry's adoption of microstructuralism as well motivated, because it is a natural 'carving' of chemical reality. I noted earlier that chemical reactions are individuated by the chemical substances with which they begin and end. Putting these two claims together, it follows that chemical reactions are the

⁴An additional argument is that if thermodynamic properties are what make a substance what it is, they should be metaphysically necessary. This is not, I believe, the perspective of chemistry, which allows thermodynamic properties such as boiling point to vary across nomologically different worlds (see Hendry, 2023; Hendry & Rowbottom, 2009).

particular reactions they are in virtue of the structures at the molecular scale with which they begin and end.

Before we move on to considering reaction mechanisms, it is worthwhile saying something about the scope and content of microstructuralism about chemical substances. On the subject of content, I disagree with the claim, made for instance by Needham (2002, 208) and Jaap van Brakel (2000, Chapter 4) that microstructuralism entails a reductionist view of substances. It involves only a claim about what chemical substances are made of, and therefore the resources that go into forming their structures at the molecular scale. This is consistent with molecules and substances being strongly emergent, or so I shall argue in the final section. On the subject of scope, by the term ‘chemical substances’ I mean chemically homogeneous stuffs as viewed by the discipline of chemistry. Some critics of microstructuralism see the fact that some other kinds of stuff, viewed from the perspectives of scientific disciplines and other kinds of activity which may be quite different from chemistry’s, are not best understood in microstructural terms as a criticism of microstructuralism (see for instance LaPorte, 2004; Havstad, 2018). However, microstructuralism about chemical substances does not apply to milk and wool (to take two examples), because they are chemically heterogeneous stuffs. Chemists don’t decide what counts as milk or wool, although they can expertly analyse particular samples.⁵ As I understand it, the same applies to protein classification as viewed from the perspective of the life sciences. Tom McLeish (2019) has argued that the physical processes underlying protein folding demonstrate many important connections with soft matter physics, a subdiscipline of condensed matter physics that studies (for instance) the mechanical, topological and thermodynamic properties of polymers, a focus that often demands that it abstracts away from their chemical composition, and engages with interactions at energy scales at which high-energy and quantum interactions are irrelevant. But the differing classification is not simply ‘higher-level’ or ‘more special’ disciplines defining their kind-terms more abstractly than ‘lower-level’, ‘less special’ or ‘less fundamental’ disciplines in a hierarchy of the sciences. Gil Santos, Gabriel Vallejos and Davide Vecchi (2020, 364) worry that microstructuralism in chemistry entails a reductionist view of processes in cells if it sees the causal powers of molecules as being determined only by the internal, or non-relational features of molecules (i.e. in abstraction from the environments in which they exert them). In my view microstructuralism about chemical substances is not a reductionist philosophy for cell biology for two kinds of reason. Firstly, proteins *in vivo* are rightly understood functionally, and so may not be heterogeneous chemical substances (their primary structure may vary). Given our earlier point about substances, they fall outside the scope of my earlier claim for microstructuralism. Secondly, as Santos, Vallejos and Vecchi point out, proteins participate in cellular processes which are subject to top-down constraints. It would be quite wrong to abstract away from the environment. This point is crucial, and

⁵David Knight (1995, Chapter 13) calls chemistry a ‘service science’ on account of its analytical expertise, which I think nicely captures its status with respect to milk and wool.

applies outside the cellular context: the causal powers of an entity at the molecular scale depend on the environment in which they are exerted. I will return to this point in the final section.

7.3 What Is a Reaction Mechanism?

Molecules are structured entities composed of electrons and atomic nuclei: they are ‘composed’ of electrons and nuclei in the sense that, if you take away the electrons and nuclei, there is nothing left. Chemical reactions must involve changes to molecular structures, and these must involve rearrangements of the electrons and nuclei. A proposed reaction mechanism is just a detailed proposal about how those rearrangements go. Hence a reaction mechanism will be a process involving rearrangements of nuclei and electrons that is, transfers of conserved quantities such as charge, mass and energy.

This rather high-level view is borne out by a closer look at chemists’ discussions of reaction mechanisms. In an influential textbook of theoretical organic chemistry, Edwin S. Gould defines ‘reaction mechanism’ as follows:

In the ideal case, we may consider the mechanism of a chemical reaction as a hypothetical motion picture of the participating atoms. Such a picture would presumably begin at some time before the reacting species approach each other, then go on to record the continuous paths of the atoms (and their electrons) during the reaction, and come to an end after the products have emerged. (Gould, 1959, 127)

Gould goes on to say that such a ‘hypothetical motion picture’ is not directly empirically accessible, and that in practice chemists focus on particular steps in the reaction:

Since it is not generally possible to obtain such an intimate picture, the investigation of a mechanism has come to mean obtaining information that can furnish a picture of the participating species at one or more crucial instants during the course of the reaction. (1959, 127)

These ‘crucial instants’ involve the making and breaking of chemical bonds, and other kinds of change that affect nuclear positions and electron distribution within the molecule.

Drawing on Gould and other textbook discussions, William Goodwin (2012) identifies two conceptions of a reaction mechanism. On the thick conception, a reaction mechanism is ‘roughly, a complete characterization of the dynamic process of transforming a set of reactant molecules into a set of product molecules’ (2012, 310). As Goodwin notes this is something like Gould’s motion picture (2012, 310). On the thin conception in contrast, mechanisms are ‘discrete characterizations of a transformation as a sequence of steps’ (2012, 310). This distinction raises three important issues: first, what relationship there is between thick and thin reaction mechanisms; second, how they are related to theories and models in chemistry, and third, how they relate to philosophical accounts of mechanisms and causation (see Goodwin, 2012, 310–15).

Felix Carroll argues that Gould's motion picture 'should be viewed as animation or simulation' because 'we cannot see the molecular events; we can only depict what we infer them to be' (1998, 317). According to Goodwin, thick mechanisms are in some sense more fundamental, perhaps because they are descriptively complete, yet they are only indirectly accessible to experiment. If they have a role in chemical thinking it is as a regulative ideal of complete description. The steps in a thin mechanism, in contrast, are more accessible to chemical inference because they affect the kinetics of the reaction (how fast the reaction goes, and how rate depends on background conditions), and they leave traces on the structures of the products. I will examine some examples later. Clearly, there is a close relationship between the two: a thick mechanism is a fuller description, but the steps in a thin mechanism can be found in a thick mechanism. For that reason, and also because thick and thin conceptions represent the very same processes, I think that even if we can draw contrasts between the two conceptions of mechanism, they ought to represent the underlying processes consistently. A thin mechanism may have gaps, but we should not think of it as representing mechanisms *as gappy*. To get to a thin mechanism from a thick mechanism we simply focus on some crucial steps, ignoring the rest. We are abstracting rather than falsifying or removing any important features of the mechanism.

Turning to the second issue, Goodwin sees thick mechanisms as more fundamental in a second sense, that they are more directly related to fundamental theory, which he identifies with potential energy surfaces (or free energy surfaces) that figure widely in discussions of mechanisms in theoretical organic chemistry.⁶ This thought further supports the idea that thick and thin mechanisms are commensurable: the steps of a thin mechanism can be identified as topological features on a PE surface. A slow, or rate-determining step will typically correspond to the traversal of a maximum.

Turning now to the last question, Goodwin (2012, 326) points out that mechanisms on the thick conception are continuous processes involving transfers of conserved quantities (for the most part, mass, charge and energy), exemplifying Wesley Salmon's processual theory of causal explanation (Salmon, 1984). In contrast, mechanisms on the thin conception, he argues, are a better fit with Peter Machamer, Lindley Darden and Carl Craver's definition of mechanisms as 'entities and activities organized such that they are productive of regular changes from start to termination condition' (2000, 3):

[M]echanisms in the thin sense are decompositions of a transformation into standardized steps, each of which is characterizable in terms of certain types of entities (nucleophiles and

⁶Potential energy (PE) surfaces aren't quite fundamental. For the evolution of a physical system to be describable in terms of a PE surface, the energy of the system must be a function of just the nuclear coordinates. As we shall see in the final section of this paper, this is a substantive physical condition (adiabatic separability of electronic and nuclear motions), and one that quantum-mechanical systems can only approximate. In general, quantum-mechanical systems do not even approximate it.

core atoms, for example) and their activities or capacities (such as capacities to withdraw electrons or hinder backside attack). (Goodwin, 2012, 326)

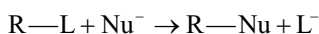
I agree with Goodwin that thin reaction mechanisms can be understood in terms of Machamer, Darden and Craver's definition, but think that their highly abstract characterisation applies just as well to thick reaction mechanisms. Conversely, thin mechanisms focus on key steps in chemical reactions, involving the making and breaking of bonds, and relative movements of nuclei. They too must involve transfers of conserved quantities. Given my earlier constraint about consistency of philosophical characterisation between the two conceptions, it seems to me that this had better be the case. Craver and Tabery (2016, 2.3.1) present no real objections to transference theories as a way of understanding mechanisms, commenting only that references to transfers of conserved quantities are rarely explicit in the special sciences and that transference theories face traditional objections concerning absence and prevention. As we have seen, references to transfers of conserved quantities are rather explicit in chemistry, and the other issues can be set aside if we regard transference as providing an account only of reaction mechanisms in chemistry, but not necessarily of causal claims more generally.

The historical development of chemists' theoretical understanding of reaction mechanisms supports the view I am defending here. From the 1860s, chemists developed a theory of structure for organic substances: how particular kinds of atoms are linked together in the molecules that characterise a substance. This theory was based on chemical evidence alone: that is, the details of which substances can be transformed into which other substances (see Rocke, 2010). Then G.N. Lewis proposed that the links between atoms in molecular structures were formed by the sharing of electrons in covalent bonds (Lewis, 1916). From the 1920s onwards, C.K. Ingold and others used these insights to develop a theory of reaction mechanisms, in which transformations between organic substances were understood as involving series of structural changes falling into a few basic kinds (Brock, 1992, Chapter 14; Goodwin, 2007). The key idea was that the making and breaking of chemical bonds should be understood in terms of the movement of electrons within the molecule. Since then, mechanisms have been central to explanations in organic chemistry, and have been fully integrated with theories of structure, with molecular quantum mechanics, and with kinetic, structural and spectroscopic evidence. Although Lewis' covalent bond may look quaint since the arrival of quantum mechanics, it was highly influential in the developing understanding of reaction mechanisms and their explanatory relationship to chemical kinetics and structure. It commensurated bonds and electrons, allowing their interconversion in descriptions of reaction processes, a key insight that we will see illustrated shortly.

As mentioned, the steps in thin reaction mechanisms fall into a relatively small number of basic kinds, each of which is well understood: an atom or group of atoms leaving or joining a molecule, and a molecule, molecular fragment or ion rearranging itself. It is the thin conception that underwrites explanations within chemical kinetics, the study of how quickly chemical reactions happen, and what determines how quickly they proceed. This is because a reaction can only proceed as fast as its

slowest step—the rate-determining step—and the rate will tend to depend only on the availability of species involved in this step.⁷

I will illustrate these points with a couple of examples. Consider an organic compound of the form R—L in which a substituent L (e.g. a halogen atom such as chlorine, bromine or iodine, or a group of atoms) is attached to a saturated hydrocarbon group R. The ‘leaving group’ L can be replaced with a nucleophilic species Nu⁻ which might be the hydroxyl ion OH⁻, or another halide ion:

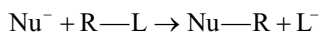


This nucleophilic substitution may happen via two different mechanisms, depending on the nature of the alkyl group R, the nature of the leaving group and the reaction conditions, including the solvent. The structures of the products can be subtly different, too. During the 1920s and 1930s Ingold developed two models (S_N1 and S_N2) of how these reactions might occur, in order to explain the contrasting chemical behaviour of different alkyl halides. In the S_N1 mechanism, the alkyl halide first dissociates (slowly) into a carbocation (or carbonium ion, in older terminology) R⁺ and the leaving group X⁻. The carbocation then combines (quickly) with the nucleophile.

- Step 1 (slow): R—L ⇌ R⁺ + L⁻
- Step 2 (fast): R⁺ + Nu⁻ → R—Nu

The slowest (and therefore rate-determining) step involves just one type of molecule, RL, and S_N1 means ‘unimolecular nucleophilic substitution’. The rate of the reaction can be expected to depend on the concentration of RL (written [RL]) and be independent of the nucleophile concentration [Nu⁻].

In the S_N2 mechanism, substitution occurs when, as described by Gould (1959, 252), the leaving group is ‘pushed off’ the molecule by the incoming nucleophile Nu⁻:



The reaction requires a bimolecular collision between the nucleophile and the target molecule (hence ‘S_N2’), so the reaction rate can be expected to be proportional to both [RL] and [Nu⁻].

Aside from the kinetics of these reactions, the S_N1 and S_N2 models also explain the differing stereochemical effects of nucleophilic substitution on different kinds of alkyl halide. Consider again the S_N1 mechanism. The molecular geometry of a saturated carbon atom is tetrahedral, and if it is bonded to four different functional groups of atoms it is asymmetrical: like a left or right hand it will not be superimposable on its mirror image (its enantiomer). In contrast the carbocation intermediate produced by step 1 has a trigonal planar geometry. So, when the nucleophile

⁷Note that this is only a tendency: if one of the reactants in another step is scarce enough in the local environment, then that step will become the slowest step.

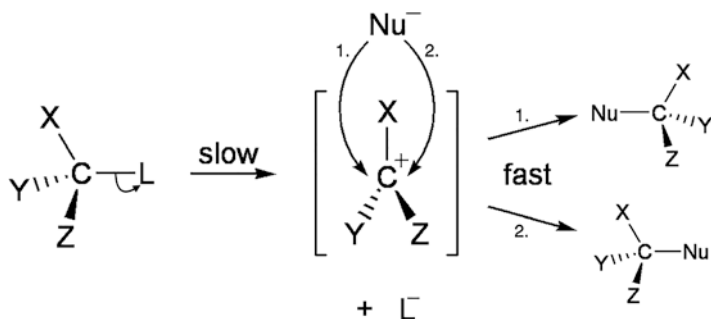


Fig. 7.2 Two possible reactions of a trigonal planar carbonium ion with nucleophile Nu^-

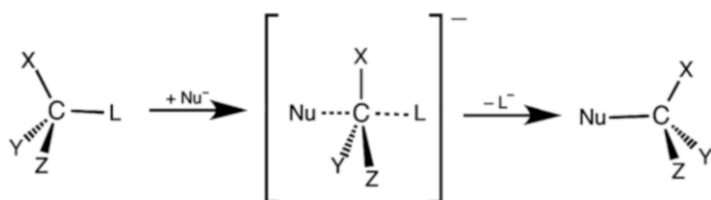
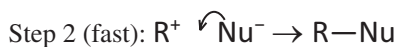
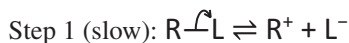


Fig. 7.3 Nucleophilic attack via the $\text{S}_{\text{N}}2$ mechanism, showing inversion of configuration

Nu^- approaches, it may do so in either of two directions, yielding an equal (racemic) mixture of two enantiomers (see Fig. 7.2).

The $\text{S}_{\text{N}}2$ mechanism, in contrast, requires a bimolecular reaction between the nucleophile and the alkyl halide. One might think that the nucleophile could approach from either the same side as the leaving group, or from the opposite side (see Fig. 7.3). In practice, ‘ $\text{S}_{\text{N}}2$ reactions invariably proceed with inversion of configuration via back-side attack.’ (Roberts & Caserio, 1965, 297). Thus in a chiral molecule the nucleophile is not simply substituted for the leaving group: the product will correspond to the enantiomer of the reagent, rather than the reagent itself.⁸

Groups of atoms leaving or joining a molecule necessarily involve the breaking or formation of bonds, and the breaking or formation of bonds involves transfer of electronic charge. This is quite explicit when curly arrows are added to the chemical equations constituting a mechanism.⁹ The $\text{S}_{\text{N}}1$ mechanism, for instance, might be represented as follows:



⁸This explains the Walden inversion, known to chemists since the 1890s (see Brock, 1992, 544–5).

⁹‘Curly arrow’ is the official name among chemists. For a discussion of their interpretation see Suckling et al. (1978, Section 4.3.3), who argue that curly arrows mean *formal* transfer of electrons. I take it that that means that the net difference between the starting and end structures is equivalent to the charge transfer indicated by the arrow.

The curly arrow in step 1 represents the net transfer of the electron pair constituting a bond to the leaving group L. Since the bond was constituted by a *shared* electron pair, one of which had been contributed by L itself, the result is a single excess negative charge on L. Hence it is a negative *ion* L^- that leaves. The curly arrow in step 2 represents the net transfer of two electrons from the nucleophile Nu^- to a bond between it and the central carbon atom in the carbocation R^+ .

The mechanism therefore implements the relationship of constitution between electron pairs and chemical bonds introduced by Lewis (1916). Realisation is also an appropriate way to characterise the relationship, however. The concept of a chemical bond was introduced into organic chemistry during the 1860s to account for the sameness and difference of various organic substances, in terms of a bonding relation holding between atoms (Rocke, 2010). In a 1936 presidential address to the Chemical Society (later to become the Royal Society of Chemistry), Nevil Sidgwick pointed out that in this structural theory ‘No assumption whatever is made as to the mechanism of this linkage.’ (Sidgwick, 1936, 533) In short, in the 1860s the organic chemists identified a theoretical role for which, in the 1910s, Lewis identified shared electron pairs as the realiser, although he did not have a detailed account (and certainly no mathematical theory) of *how* electrons did the realising. The relationship remains complicated: the central explanatory theory is quantum mechanics, but that theory itself, applied to ensembles of electrons and nuclei, has nothing to say on the subject of bonds. Chemists need to *find* the bonds in quantum mechanics: the theory can, with the help of the Born-Oppenheimer approximation (localisation of the nuclei, and instantaneous neglect of their motion) provide an excellent theory of the electron-density distribution and how it interacts with changing nuclear positions. Mathematical analysis of a molecule’s electron-density distribution yields bond paths between atoms which resemble the familiar bond topologies of chemistry (see Bader, 1990; Popelier, 2000).

Reaction mechanisms are *possible* pathways from reagents to products in two distinct senses of possibility. In the case of S_N1 and S_N2 , both pathways might be physically possible even if one will typically be favoured. Both might be followed by different molecules in the same reaction vessel. Both, in fact, are actual pathways. However, Roberts and Caserio discuss two possible routes for S_N2 reactions, involving front-side and back-side approach, but note that the back-side approach ‘invariably’ occurs (Roberts & Caserio, 1965, 297). The front-side approach is (in some weak sense) geometrically possible, but is rendered physically *impossible* by the structure of the species being attacked. One might also say that, although the experimental evidence seems to rule out front-side attack for S_N2 reactions, until that information was acquired it might have been considered *epistemically* possible. This suggests a role for eliminative reasoning about reaction mechanisms, something that Roald Hoffman (1995, Chapter 29), illustrates beautifully with a discussion of three possible mechanisms for the photolysis of ethane to ethene (known traditionally as ethylene), and how H. Okabe and J. R. McNesby used isotopic labelling to eliminate two of them. In photolysis, light energy (written $h\nu$) causes ethane (C_2H_6 or H_3C-CH_3) to eliminate hydrogen (H_2), leaving ethene (a molecule with a carbon-carbon double bond; see Fig. 7.4).

Fig. 7.4 Photolysis of ethane: the reaction. (From Hoffmann, 1995, 145)

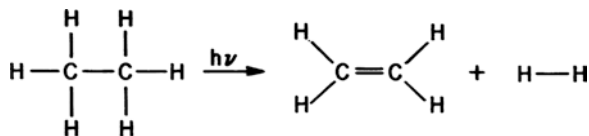
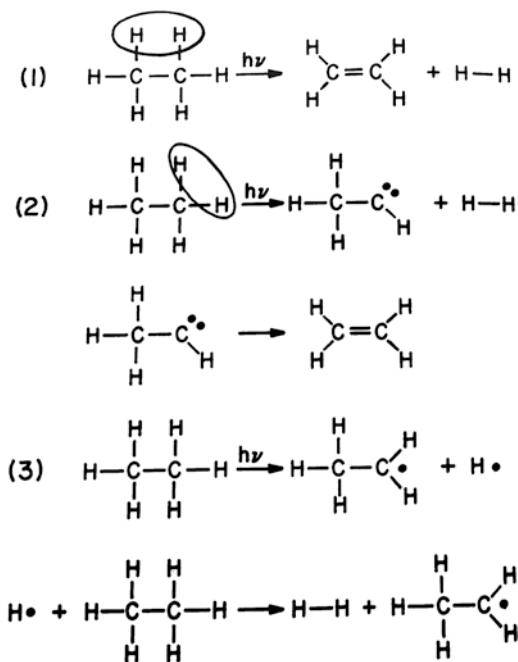


Fig. 7.5 Photolysis of ethane: three possible mechanisms. (From Hoffmann, 1995, 146)



The question is how this occurs, with three candidate pathways (see Fig. 7.5). Mechanism 1 involves two hydrogen atoms on neighbouring carbon atoms leaving by a ‘concerted reaction’. Mechanism 2 involves two steps: elimination of H₂ from a single carbon atom followed by rearrangement of the fragment to give ethene. In mechanism 3, a light photon breaks a C-H bond, leaving two free radicals (C₂H₅• and H•, the dots representing unpaired electrons). This kicks off a chain reaction when the free radicals collide with other molecules, generating further free radicals. Mechanism 3 is eliminated because in a mixture of C₂H₆ and C₂D₆ (deuterated ethane), it would produce significant amounts of HD (isotopically mixed hydrogen molecules), whereas little is detected. That leaves mechanisms 1 and 2, of which mechanism 1 is eliminated because it would produce significant amounts of HD from isotopically mixed ethane (H₃C—CD₃), whereas little is detected.

That leaves mechanism 2: does the elimination of the other candidate mechanisms afford it any positive evidential support? Chemists often say that mechanisms cannot be proven (see for instance Sykes, 1981, 43; Carpenter, 1984, Chapter 1). Hoffmann notes the Popperian response but notes that scientists ‘want to do something positive’ (1995, 149). I think Hoffmann is right here, and the right framework

for that is eliminative induction rather than Popperian refutation. For philosophers of science, the interest is whether structure theory (and knowledge of other mechanisms) can do any epistemic heavy lifting here, by limiting the number of conceivable pathways. Eliminative inductions, notoriously, are only as secure as the initial disjunctions of possibilities that constitute their major premises. A well-attested structure for the reagent will constrain how it might possibly transform into the product. The epistemic justification for the attribution of a structure is surely the source of any justification for the starting disjunction.

7.4 Mechanisms and Reduction¹⁰

Many philosophers would expect the view I am defending of mechanisms in organic chemistry to support a robustly reductionist view of chemical substances. Firstly, it is committed to microstructural essentialism about chemical substances, which supports theoretical identity claims of the form ‘gold is the element with atomic number 79’ and ‘water is H₂O’. Secondly, I have defended the claim that chemical reaction mechanisms involve transfers of conserved quantities, primarily charge and mass, in the form of electrons and nuclei. Isn’t it obvious that if water is H₂O, then everything that water does is done by entities at the molecular level, namely H₂O molecules? And isn’t the path to reductionism even clearer once we realise that the mechanisms by which water exerts its powers – *how* it does what it does – are properly described in terms of processes involving transfers of physical quantities governed by physical laws?

I think both inferences should be rejected. Microstructuralism about chemical substances is compatible with their being strongly emergent. Hilary Putnam once said that the extension of ‘water’ is ‘the set of all wholes consisting of H₂O molecules’ (1975, 224). That is the first mistake, if a ‘whole’ is taken to be a mereological sum, or any other whole that can be formed without any interaction among the parts. ‘Water is H₂O’ is true, but that doesn’t mean that H₂O molecules are all there is to water. Water is formed from H₂O molecules interacting: some of them self-ionise, producing protons (H⁺) and hydroxyl ions (OH⁻). Others form hydrogen-bonded oligomolecular structures. If Putnam had been right, then it would be safe to assume that water could not have any causal powers over and above those inherited from its constituent H₂O molecules. Such a whole has no bulk properties, so there is no distinction to be made between its molecular and its bulk properties. In contrast steam, liquid water and ice (which has various structures) do have distinct properties produced by the distinct kinds of interactions between their parts.

¹⁰This section draws on published research arising from the Durham Emergence Project. The first part draws on arguments from Hendry (2017a, b); the second part draws on collaborative work with Robert Schoonmaker, which I have presented in Hendry (2019, 2022). I am most grateful to the John Templeton Foundation for funding the project (Grant ID 40485), and also to members of the project for many helpful conversations.

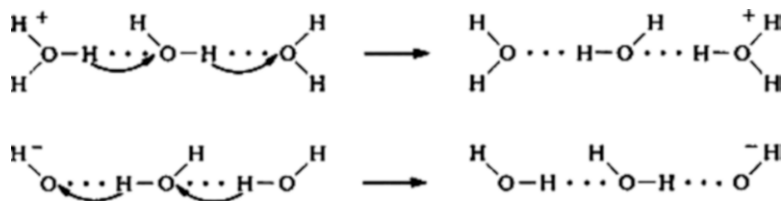


Fig. 7.6 Mechanisms of proton and hydroxyl ion transfer in water. (From Eisenberg & Kauzmann, 1969, 227)

Wherever there is significant interaction between the H₂O molecules, there is scope for that interaction to bring new powers into being. This is particularly clear if that interaction includes self-ionisation and the formation of oligomers. Now to mechanisms: as Eisenberg and Kauzmann put it, protons and hydroxyl ions both have ‘abnormal mobilities in both ice and water’ (1969, 226), which is possible because their excess charge can travel along the hydrogen bonded supramolecular structures without matter having to make the journey to carry them (see Fig. 7.6).

The mechanism by which that power is exercised requires some part of the molecular population to be charged. It also requires the supramolecular organisation. It therefore depends on a feature of a diverse *population* of molecular species. The reductionist will say at this point that the water can only acquire its causal powers from its parts, and interactions between them. Therefore, no novel causal powers have been introduced. The strong emergentist will ask why, when it is being decided whether they are novel, the powers acquired only when the molecules interact are already accounted for by the powers of H₂O molecules. If the reductionists’ claim is that any power possessed by any molecular population produced by any interaction between H₂O molecules is included, and we know that independently of any empirical information we might ever acquire about what water can do and how it does it, then it seems that we know *a priori* that there will be no novel causal powers, and therefore that reductionism is true. I take it that this would be for the reductionist to beg the question. This does not of course mean that the strong emergentist wins the argument by default: only that in making their case reductionists should deploy topic-specific scientific arguments. In the absence of such arguments, the reductionist and the strong emergentist conclude this discussion honours even. Anti-reductionists need not fear theoretical identities, and should even learn to love them.¹¹

Reductionists here point out that quantum mechanics provides an account of the structure of molecules. I agree, but a detailed examination of *how* molecular structures are in fact explained within quantum mechanics puts pressure on the idea that they can be said to be reduced to quantum mechanics. Quantum-mechanical explanations of structure depend on non-trivial assumptions about physical interactions between the electrons and nuclei within a molecule. Emergence provides a fruitful

¹¹ For similar reasons I have never understood why ‘pain is c-fibres firing’ should have the significance for the reductionism debate that it has. The identification would only work from an explanatory point of view by appealing to significant levels of organisation above the level of the c-fibres.

and flexible framework within which to think about these assumptions: it is well known that there are different kinds of emergence (e.g. strong *vs.* weak; ontological *vs.* epistemological), but for the purposes of this section it will be helpful to set aside the issue of which particular species of emergence is at stake. My strategy will be to identify the additional assumptions required to explain structure, and then argue that they fall under a widely accepted abstract characterization of emergence as dependent novelty.¹²

Quantum mechanics is generally understood to describe many-body systems of electrons and nuclei in terms of the Schrödinger equation. The idea is that we will seek solutions to the Schrödinger equation that correspond to the possible stationary states of the molecular system. Electronic and nuclear motions are first separated, on account of the very different rates at which they move and respond to external interactions. This is the adiabatic approximation, which yields wavefunctions for two coupled systems (of electrons, and of nuclei), dynamically evolving in lockstep. Essentially the same assumption is widely referred to as the ‘Born-Oppenheimer’ approximation and justified solely in terms of the difference between the masses of electrons and nuclei.¹³ That cannot be a sufficient justification, however, because the ratio of the nuclear and electronic masses is constant, but the adiabatic approximation breaks down in many systems. So what is the difference between adiabatic and non-adiabatic cases? Not the ratios of the nuclear and electronic masses, but rather the timescales over which electronic and nuclear wavefunctions are assumed to respond to each other. The justification for the adiabatic approximation should therefore be understood in terms of timescales too. Consider a single electron in a one-dimensional box: it is often observed that an adiabatic system (i.e. one in which the adiabatic approximation holds) is one in which the walls of the box move slowly enough for the electron’s wavefunction to respond smoothly and continuously to the change. If the walls of the box move too quickly, the system will jump to a different quantum state. Likewise, the adiabatic approximation allows us to think of the joint electronic and several nuclear wavefunctions as fixed parameters with respect to each other, each acting with respect to the other like the walls of the one-dimensional box on the electron’s wavefunction. The electrons see the nuclei as static, while the nuclei see the electrons as smeared-out charge distribution.

When modelling a molecular structure, the nuclei are assigned positions corresponding to the equilibrium positions of the known structure. Density-functional theory (DFT), which has revolutionised molecular quantum mechanics in the last few decades (see Kohn, 1999), then replaces the $3N$ -dimensional electronic wavefunction with a 3-dimensional electron density function: it can be shown that this can be done without approximation. According to the Hellmann-Feynman theorem, the overall force on a nucleus in the system is determined by the electron density, so

¹²I am most grateful to Stewart Clark, Tom Lancaster and Robert Schoonmaker for conversations on these topics. This section draws on joint work with Schoonmaker, but the position I set out is my own interpretation of the various scientific facts, and I would not wish to implicate these interlocutors in any of my misunderstandings.

¹³See for instance Atkins (1986, 375).

effectively the nuclei are being pushed around by their interactions with the electrons.¹⁴ As Richard Bader notes, this is quite intuitive:

Accepting the quantum mechanical expression for the distribution of electronic charge as given by $\rho(r)$, the theorem is a statement of classical electrostatics and therein lies both its appeal and usefulness. (Bader, 1990, 315)

That a particular substance has a particular molecular structure is then explained by showing that that structure corresponds to a configuration in which the energy is a minimum (i.e. the forces on the nuclei are effectively zero). Two points are worth making, concerning scope and explanatory power. First consider scope: the Schrödinger equation for a molecule depends only on the electrons and nuclei present. Hence different isomers share their molecular Schrödinger equations, the starting point of the above explanation. Ethanol ($\text{CH}_3\text{CH}_2\text{OH}$) and dimethyl ether (CH_3OCH_3) share the same Schrödinger equation, as do enantiomers such as L- and D-tartaric acid (see Sutcliffe & Woolley, 2012). The starting point of the explanation – the molecular Schrödinger equation – does not respect the differences between isomers, and what results from localising the nuclei within the adiabatic approximation *does* respect these differences, so localising the nuclei in positions corresponding to the different isomers effectively inserts these differences. Reductionists may see this move as having a pragmatic justification: we cannot directly solve the molecular Schrödinger equations, and so must introduce approximations. But it is hard to see the Born-Oppenheimer approximation as a mere approximation if it changes the scope of the quantum-mechanical description, making it apply only to one isomer rather than to all of them.

A second point concerns explanatory power. Approximations are widely assumed to have no independent explanatory power because they are proxies for exact equations: anything that could be explained using an approximate or idealised model could, in principle, be explained using the exact equations. This would be a reasonable thing to conclude if every explanatorily relevant feature of the model could be grounded in the exact equations in some way. In my first paper on these topics (Hendry, 1998) I argued, drawing on the work of Brian Sutcliffe and Guy Woolley, that defences of Born-Oppenheimer models based on the proxy view must fail because they change symmetry properties, which are explanatorily relevant features with respect to which isomers may differ, and cannot therefore be grounded in the molecular Schrödinger equation alone. It seems hard to argue that the approximations have no independent explanatory power. The molecular structures could not, *in principle*, be explained without them. Hence the different structures, their different symmetry properties and the different causal powers they ground, are effectively introduced as unexplained explainers unless we regard the adiabatic separability of the nuclear and electronic motions and the localization of the nuclei as part of the explanation.

¹⁴The quantum-mechanical electron-density distribution carries information about the nuclei too, so each of the nuclei is really being ‘pushed around’ by its interactions with the entire system.

One might regard these conditions merely as initial or boundary conditions, and no more interesting than the ‘auxiliary assumptions’ that, according to received wisdom in the philosophy of science ever since Duhem, are required when we apply any theory. The story goes something like this: quantum mechanics (QM) implies the existence of molecular structure (MS) only in conjunction with statements describing the necessary boundary and initial conditions (BIC). Thus the conjunction of QM and BIC implies MS. This is all correct: adiabatic separability and nuclear localisation might plausibly be thought of as boundary conditions, while the choice of nuclear positions looks like an initial condition. However, this is no help at all if we want to see this explanation as a derivation from quantum mechanics, because meeting the adiabaticity and nuclear localisation conditions exactly is impossible for any genuine quantum system. The conjunction of QM and BIC is, in some important sense, incoherent.

I therefore think it is much less puzzling to describe the situation as follows: the mathematics provides only what one might call a dynamical consistency proof: the conditions that define a Born-Oppenheimer model *could not* hold exactly in any fully quantum-mechanical system, but the two kinds of system will evolve dynamically in approximately similar ways, for some given level of accuracy, and over the timescales relevant to the calculation. As we have seen, the molecular structure calculations described above assume dynamical conditions—the adiabatic separability of electronic and nuclear motions, and the localisation of the nuclei—which *could not* hold exactly in any quantum system. All that can be concluded therefore is that a quantum-mechanical system of electrons and nuclei will display approximately similar dynamics to the model. No derivation of the model dynamics from the exact equations has been provided, nor even a demonstration of their consistency under the conditions. All that the mathematics provides is that the relevant approximations introduced in the model can be neglected for a given level of accuracy over relevant timescales.

In joint work with Robert Schoonmaker (Hendry & Schoonmaker, [Forthcoming](#)), rather than treating adiabatic separability and nuclear localisation as approximations, we interpret them as substantive special assumptions about dynamical interactions within a quantum-mechanical system of electrons and nuclei. As already noted, these conditions radically transform the dynamical behaviour of quantum systems, and the scope of the equations that describe them. Adiabatic separability makes the overall energy of the electrons and nuclei a function of the nuclear configuration, so that the dependence of energy on nuclear positions can be mapped by a potential energy (PE) surface (or rather a hypersurface). This is not a global assumption because it depends on adiabatic separability, and a system of electrons and nuclei will not have a global PE surface. PE surfaces are not foliated, and near where they cross, the adiabatic separability of nuclear and electronic motions breaks down (see Lewars, 2011: Chapter 2). The effect of nuclear localisation is just as radical and interesting, for it suppresses the dynamical expression of quantum statistics. In general, any quantum system of electrons and nuclei must obey nuclear permutation symmetries: the overall wavefunction must be symmetric (for bosons) or antisymmetric (for fermions). These symmetries correspond to real physical

processes however: a molecule exploring the space of its possible nuclear permutations involves the exchange of identical particles. Physical conditions that tend to slow down the exchange processes allow the particle permutation symmetries to be neglected over timescales that are relatively short compared to the exchange. In a quantum system with a classical molecular structure, the nuclei can typically be regarded as being localised by their interaction with the rest of the system. The dynamical effect is that which kind of statistics are assumed to apply to the nuclei – whether the overall wavefunction is symmetric (Bose-Einstein statistics), anti-symmetric (Fermi-Dirac statistics) or indeed *asymmetric* (classical statistics) with respect to permutation of the nuclei – makes a negligible difference to the evolution of the system. Interaction with the rest of the system effectively transforms the nuclei from quantum entities into classical objects.

It should be emphasised that neither of these conditions is necessary for bonding as such: chemists and condensed matter physicists study systems such as metals and superconductors in which there is bonding (since they form cohesive materials), but in which nuclear and electronic motions are not adiabatically separable, and in which the nuclei are not localised, in the sense that quantum statistics must be taken into account in describing their structure and behaviour. These conditions should be regarded as necessary only for the kind of structure that is describable in terms of the classical chemical structures developed in organic chemistry during the nineteenth century, and some later generalizations. Interestingly, although the expression of nuclear permutation symmetries is generally suppressed, there are molecules, such as protonated methane, in which interactions between one pair of protons means that the symmetries are expressed in the dynamical behaviour of the molecule (see Marx & Parrinello, 1995; Hendry & Schoonmaker, [Forthcoming](#)). The above conditions—adiabatic separability of nuclear and electronic motions, and nuclear localisation—are not, moreover, sufficient for the emergence of classical molecular structure. Some molecules, such as cyclobutadiene, tunnel between two different structures each of which is expressed in the molecule's interaction with radiation: in IR spectra the molecule exhibits square symmetry, while higher-frequency x-ray diffraction catches it in the rectangular states between which it tunnels (see Schoonmaker et al., 2018).¹⁵ It should be emphasised that tunnelling between different classical structures is the normal quantum-mechanical behaviour (consider P.W. Anderson (1972) on ammonia), but the dynamical behaviour of many molecules can be understood in terms of a single classical structure. Hence dynamical restriction to a single structure is a third necessary condition for the classical kind of structure that is exhibited by many organic molecules and was discovered by organic chemists in the 1860s.

In my view these considerations provide a good argument for regarding adiabatic separability, nuclear localization and dynamical restriction to a single structure as substantive conditions that form a necessary part of the explanation of this kind of

¹⁵This is an expression of the scale-relativity of structure, for which I have argued elsewhere (see Hendry, 2023).

structure. In what sense should structure be regarded as emergent, however? Emergence is often understood as dependent novelty: emergent properties are borne by systems that depend for their existence on something more fundamental (typically their parts), but also display properties or behaviour that is in some significant way novel with respect to the parts. This applies readily to the foregoing discussion. Molecular structures are ontologically dependent on electrons and nuclei: they cannot exist without them. The novelty consists in the distinct dynamical behaviour displayed by electrons and nuclei in the context of structured systems: adiabatic separability, nuclear localization and restriction to a single classical structure, which in each case is a suspension of the normal behaviour of a quantum system. The adiabatic separability and localization are also examples of *transformational* emergence (see Santos, 2015; Humphreys, 2016, Chapter 2), in which the behaviour of the parts of an emergent system is so different that it makes sense to say that they have been transformed into a new kind of entity. The radical transformation of nuclei from entities that obey quantum statistics into localized, semi-classical entities would seem to be a good example of transformational emergence.

References

- Anderson, P. W. (1972). More is different. *Science*, 177, 393–396.
- Atkins, P. W. (1986). *Physical chemistry* (Third ed.). Oxford University Press.
- Bader, R. F. W. (1990). *Atoms in molecules: A quantum theory*. Oxford University Press.
- Brock, W. H. (1992). *The Fontana history of chemistry*. Fontana Press.
- Carpenter, B. (1984). *Determination of organic reaction mechanisms*. Wiley.
- Carroll, F. A. (1998). *Perspectives on structure and mechanism in organic chemistry*. Brooks/Cole.
- Craver, C., & Tabery, J. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2016 ed.) <http://plato.stanford.edu/archives/spr2016/entries/science-mechanisms/>
- Eisenberg, D., & Kauzmann, W. (1969). *The structure and properties of water*. Oxford University Press.
- Goodwin, W. M. (2007). Scientific understanding after the Ingold revolution in organic chemistry. *Philosophy of Science*, 74, 386–408.
- Goodwin, W. M. (2012). Mechanisms and chemical reaction. In R. F. Hendry, P. Needham, & A. I. Woody (Eds.), *Handbook of the philosophy of science, volume 6: Philosophy of chemistry* (pp. 309–327). North-Holland.
- Gould, E. S. (1959). *Mechanism and structure in organic chemistry*. Holt, Rinehart and Wilson.
- Havstad, J. C. (2018). Messy chemical kinds. *British Journal for the Philosophy of Science*, 69, 719–743.
- Hendry, R. F. (1998). Models and approximations in quantum chemistry. In N. Shanks (Ed.), *Idealization in contemporary physics: Poznan studies in the philosophy of the sciences and the humanities* 63 (pp. 123–142). Rodopi.
- Hendry, R. F. (2006). Elements, compounds and other chemical kinds. *Philosophy of Science*, 73, 864–875.
- Hendry, R. F. (2010a). The elements and conceptual change. In H. Beebe & N. Sabbarton-Leary (Eds.), *The semantics and metaphysics of natural kinds* (pp. 137–158). Routledge.
- Hendry, R. F. (2010b). Entropy and chemical substance. *Philosophy of Science*, 77, 921–932.

- Hendry, R. F. (2017a). Prospects for strong emergence in chemistry. In M. Paolini Paoletti & F. Orilia (Eds.), *Philosophical and scientific perspectives on downward causation* (pp. 146–163). Routledge.
- Hendry, R. F. (2017b). Mechanisms and reduction in organic chemistry. In M. Massimi, J. W. Romeijn, & G. Schurz (Eds.), *EPSA15 selected papers: The 5th conference of the European Philosophy of Science Association in Düsseldorf* (pp. 111–124). Springer.
- Hendry, R. F. (2019). Emergence in chemistry: Substance and structure. In S. C. Gibb, R. F. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 339–351). Routledge.
- Hendry, R. F. (2022). Quantum mechanics and molecular structure. In O. Lombardi et al. (Eds.), *Philosophical perspectives on quantum chemistry* (pp. 147–171). Springer.
- Hendry, R. F. (2023). Structure, essence and existence in chemistry. *Ratio*, 36. In press, forthcoming.
- Hendry, R. F. (Forthcoming). *How (not) to argue for microstructural essentialism*.
- Hendry, R. F., & Rowbottom, D. P. (2009). Dispositional essentialism and the necessity of laws. *Analysis*, 69, 668–677.
- Hendry, R. F., & Schoonmaker, R. (Forthcoming). *The emergence of the chemical bond*. Unpublished manuscript.
- Hoffmann, R. (1995). *The same and not the same*. Columbia University Press.
- Humphreys, P. (2016). *Emergence: A philosophical account*. Oxford University Press.
- Knight, D. M. (1995). *Ideas in chemistry: A history of the science*. Athlone.
- Kohn, W. (1999). Electronic structure of matter: Wave functions and density functionals. *Reviews of Modern Physics*, 71, 1253–1266.
- Kragh, H. (2000). Conceptual changes in chemistry: The notion of a chemical element, ca. 1900–1925. *Studies in History and Philosophy of Modern Physics*, 31B, 435–450.
- LaPorte, J. (2004). *Natural kinds and conceptual change*. Cambridge University Press.
- Leigh, G. J., Favre, H. A., & Metanomski, W. V. (1998). *Principles of chemical nomenclature: A guide to IUPAC recommendations*. Blackwell Science.
- Lewars, E. (2011). *Computational chemistry: Introduction to the theory and applications of molecular and quantum mechanics* (Second ed.). Springer.
- Lewis, G. N. (1916). The atom and the molecule. *Journal of the American Chemical Society*, 38, 762–785.
- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Marx, D., & Parrinello, M. (1995). Structural quantum effects and three-centre two-electron bonding in CH_3^+ . *Nature*, 375, 216–218.
- McLeish, T. (2019). Soft matter: An emergent interdisciplinary science of emergent entities. In S. C. Gibb, R. F. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 248–264). Routledge.
- Needham, P. (2002). The discovery that water is H_2O . *International Studies in the Philosophy of Science*, 16, 205–226.
- Needham, P. (2008). Is water a mixture? Bridging the distinction between physical and chemical properties. *Studies in History and Philosophy of Science*, 39, 66–77.
- Needham, P. (2011). Microessentialism: What is the argument? *Noûs*, 45, 1–21.
- Popelier, P. (2000). *Atoms in molecules: An introduction*. Pearson.
- Putnam, H. (1975). The meaning of “meaning”. In *Mind language and reality* (pp. 215–271). Cambridge University Press.
- Roberts, J. D., & Caserio, M. C. (1965). *Basic principles of organic chemistry*. W.A. Benjamin.
- Rocke, A. J. (2010). *Image and reality: Kekulé, Kopp and the scientific imagination*. Chicago University Press.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Santos, G. (2015). Ontological emergence: How is that possible? Towards a new relational ontology. *Foundations of Science*, 20, 429–446.

- Santos, G., Vallejos, G., & Vecchi, D. (2020). A relational constructionist account of protein macrostructure and function. *Foundations of Chemistry*, 22, 363–382.
- Schoonmaker, R. T., Lancaster, T., & Clark, S. (2018). Quantum mechanical tunneling in the automerization of cyclobutadiene. *Journal of Chemical Physics*, 148, 104109.
- Sidgwick, N. V. (1936). Structural chemistry. *Journal of the Chemical Society*, 149, 533–538.
- Suckling, C. J., Suckling, K. E., & Suckling, C. W. (1978). *Chemistry through models: Concepts and applications of modelling in chemical science, technology and industry*. Cambridge University Press.
- Sutcliffe, B. T., & Woolley, R. G. (2012). Atoms and molecules in classical chemistry and quantum mechanics. In R. F. Hendry, P. Needham, & A. I. Woody (Eds.), *Handbook of the philosophy of science, volume 6: Philosophy of chemistry* (pp. 387–426). North-Holland.
- Sykes, P. (1981). *A guidebook to mechanism in organic chemistry* (Fifth ed.). Longmans.
- van Brakel, J. (2000). *Philosophy of chemistry*. Leuven University Press.
- van der Vet, P. (1979). The debate between F.A. Paneth, G. von Hevesy and K. Fajans on the concept of chemical identity. *Janus*, 92, 285–303.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 8

A Commentary on Robin Hendry's Views on Molecular Structure, Emergence and Chemical Bonding



Eric Scerri

Abstract In this article I examine several related views expressed by Robin Hendry concerning molecular structure, emergence and chemical bonding. There is a long-standing problem in the philosophy of chemistry arising from the fact that molecular structure cannot be strictly derived from quantum mechanics. Two or more compounds which share a molecular formula, but which differ with respect to their structures, have identical Hamiltonian operators within the quantum mechanical formalism. As a consequence, the properties of all such isomers yield precisely the same calculated quantities such as their energies, dipole moments etc. The only means through which the difference between the isomers can be recovered is to build their structures into the quantum mechanical calculations, something that is carried out by the application of the Born-Oppenheimer approximation. Consequently, it has been argued by many authors that molecular structure is written in 'by hand' rather than derived. Robin Hendry is one such author, but he goes a great deal further by proposing that this situation implies the existence of emergence and downward causation. In the current article I argue that there are alternative explanations which render emergence and downward causation redundant. Such an alternative lies in the notion of quantum decoherence and the appeal to work in the foundations of physics, which posits that the various isomers exist as a superposition until their wavefunctions are collapsed either by observation or by interacting with their environment.

Hendry also alludes to a debate among chemists as to whether chemical bonds are real or not, in the sense of directional connections between two or more nuclei in any given molecule. I reject this view and propose that the structural and energetic views of chemical bonding, that have been discussed by some philosophers of chemistry including Hendry, do not refer to any essential ontological differences. I agree that chemists view bonding in a more realistic fashion and may consider bonds to be in some senses real, while physicists may consider bonding in more abstract energetic terms. However, I do not believe that such differences in scientific practice and attitudes should be considered to offer a window as to the ontological

E. Scerri (✉)

Department of Chemistry & Biochemistry, UCLA, Los Angeles, CA, USA

e-mail: scerri@chem.ucla.edu

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_8

161

status of bonding or whether bonding is real. Finally, I discuss the kinetic energy school of chemical bonding which would seem to challenge any notion of bonds as directional entities, since bonding is no longer regarded as being primarily due to the build-up of electron density between nuclei.

Keywords Emergence · Reduction · Born-Oppenheimer · Causation · Molecular structure · Bonding

8.1 Introduction

Over a period of many years Robin Hendry has proposed a number of related views on the philosophy of chemistry. In the present article I intend to examine some of these views in detail. Like many other philosophers of chemistry before him, Hendry has worked on the question of molecular structure and its relationship with quantum mechanics.¹ Molecular structure is of course a central and important concept in chemistry with an enormous amount of experimental evidence to support its existence. Similarly, quantum mechanics represents a major pillar of modern physics and a dominant paradigm for the study of radiation and matter, which has yet to be refuted after about 100 years since it was first developed.²

The problem lies in trying to connect molecular structure with quantum mechanics. To cite a common example that is discussed in the literature, a pair of isomers, such as C_2H_5OH (ethanol) and CH_3OCH_3 (dimethyl ether) have different molecular structures even though they share precisely the same Hamiltonian operator within their quantum mechanical description.

When the Hamiltonian operates on the wavefunction for these molecules it therefore yields precisely the same energy, as well as any other properties that one may care to extract from such computations. Briefly put, quantum mechanics appears to be incapable of distinguishing between two such isomers unless one important further step is taken, namely the introduction of the so-called Born-Oppenheimer approximation. This procedure corresponds to assuming that the positions of all the nuclei in a molecule are stationary, relative to the movement of their far lighter electrons.

As a result of this approach the act of solving the Schrödinger equation for these molecules is simplified considerably. In other words, the structure of the molecule, as defined by the positions of the nuclei, is written into solution of the problem from the outset. Quantum mechanics does not therefore derive the structure of the molecule since one assumes it from the start.

This situation is somewhat analogous to that of the old quantum theory in the early years of the twentieth century. The Bohr model was successful at describing

¹Primas (1983), Woolley (1976).

²Histories of quantum theory and the later quantum mechanics include Jammer (1966) and Mehra and Rechenberg (1987).

one-electron systems but the quantization condition for the energy of the electrons had to be written into the treatment from the beginning. Stated otherwise, quantization was assumed rather than being derived. As I see it, a similar situation exists in the molecular structure problem, in which structure is typically assumed rather than being derived.

If my proposed analogy has any validity, one may wonder whether future developments in quantum mechanics might not resolve the molecular structure issue and render structure derivable.³ So far, the story I have sketched is well known and has been addressed by many authors from different perspectives (Primas, 1983; Woolley, 1976).

What Hendry brings to this issue is the view that this 'gap' between molecular structure and quantum mechanics should be interpreted as indicating that molecular structure 'emerges' in some sense. Furthermore, Hendry proposes that we should think of two kinds of molecular Hamiltonians. First of all, he speaks of the true, or resultant, Hamiltonian which does not help itself to the Born-Oppenheimer approximation, meaning that molecular structure is not assumed from the outset. He then proceeds to contrast this form with a Hamiltonian that does make use of the B-O approximation, in which the nuclear positions are fixed and which he terms the configurational Hamiltonian. Hendry also accompanies this proposal with the radical claim for the existence of downward causation, through which molecular goings on can somehow influence their component particles.

Let us return for now to the notion of emergence. Hendry has claimed that according to the current state of quantum chemistry, there is at least as much evidence for emergence as there is for the ontological reduction of chemistry, but concludes by favoring emergence. This claim would seem to be rather extravagant, at least to the present author as I have argued in more detail in a previous publication (Scerri, 2012).⁴

8.2 On Epistemological and Ontological Reduction

Hendry quite correctly contends that quantum mechanical theory is abstract, whereas any particular situation is highly specific and necessitates the use of approximations. It is possible, he continues, that any failure of reduction can be

³One possible candidate for such a development has already been outlined by Sir Roger Penrose who believes that gravity modifies quantum mechanics in a profound manner which, among other things, may provide a natural explanation for what happens during the collapse of the wavefunction (Zurek, 2003).

⁴In any case, the burden of proof lies with those who claim the existence of emergence, rather than for critics to have to provide detailed counter arguments as to why it does not even exist. Emergence may well be a buzz word in the philosophy of science literature but there is no agreement as to how it can be characterized. It certainly has no traction among the vast majority of working scientists with the exception of some cosmologists who have argued that space-time somehow 'emerges' from more fundamental quantum levels or reality (Gambini & Pullin, 2020).

attributed to making such approximations. If so, then a reduction would have failed on epistemological or inter-theoretical grounds. One cannot conclude, Hendry argues, that there is a lack of ontological reduction. So far, I am in complete agreement.

Hendry also points out that a pair of disciplines, such as chemistry and physics, typically develop independently as history unfolds and that there is no guarantee that the two sciences should mesh together perfectly in such a way that reduction could ever be established. If this is the case, then once again any apparent lack of reduction can be attributed to inter-theoretical issues and one cannot rule out the ontological reduction of one level to another one.

However, the failure of reductionism on these sorts of grounds cannot be conclusive when it comes to the more general question of ontological reduction. In order to articulate a form of ontological reduction, we need to look elsewhere. Hendry then turns to the more difficult task of venturing an opinion concerning ontological reduction,

if the reduction debate is to develop beyond the impasse over inter-theoretic reduction, it must turn to the ontological relationships between the entities, processes, and laws studied by different sciences, which are fallibly and provisionally described by their theories. One obvious requirement on a criterion of ontological reduction is that whether or not it obtains must be a substantive metaphysical issue that transcends the question of what explanatory relationships exist between theories now, or might exist in the future, even though inter-theoretic relationships must continue to be relevant evidence (Hendry 2010, p. 184).

This is an important point that, as I believe, Hendry fails to embrace fully when he addresses the issue in more detail. Moreover, I suggest that it is rather difficult to give arguments that transcend our current explanatory schemes and theories. As I see it, Hendry and other authors who claim to separate ontological question from inter-theoretical questions by focusing on entities rather than theories, may be mistaken.

Hendry continues,

reducibility is at the strong end of the spectrum because it is the limiting case that denies the distinct existence of what is dependent—the reductionists slogan is that x is reducible to y just in case x is ‘nothing but’ its reduction base, y . One can imagine many ways to cash out this slogan, depending on the aspect under which the reduced is held to be ‘nothing but’ its reduction base, but a consensus has emerged in recent philosophy of mind that the relevant aspect should be causal. Alexander’s dictum is the principle, often cited by Kim (1998, p. 119, 2005, p. 159), according to which being real requires having causal powers (Hendry 2010, p. 184).

This appears to represent a major pivot which deserves more scrutiny, namely the connection between the question of causation and that of reduction. First of all, the fact that a consensus may have arisen in the philosophy of mind may not be relevant to research in the philosophy of chemistry. Why after all should one accept a consensus that may have emerged in a completely different branch of philosophy? Moreover, the importance of causation is far from universally accepted in the philosophy of science and indeed there is a growing belief among philosophers of physics, and others, that not all explanations are necessarily of a causal nature (Norton, 2003; Lange, 2013). In addition, some theoretical chemists have also

recently denied the notion that causation plays any role whatsoever in the domain of chemistry (Matta, 2023).

The re-appearance of causes in the philosophy of science, after they had been abolished by the Logical Positivists, is a complicated issue whose examination would take us too far afield and will not be considered here (Scerri, 2021). Suffice it to say that the symmetry between explanation and derivation which existed in the logical positivist account of science became threatened because of some cases which represented a derivation while it appeared as though there was no explanation.

The classic instance of this kind is one concerning a flagpole and the shadow that it casts on a sunny day. One can calculate the length of the shadow from the height of the pole and a little trigonometry. Conversely one can calculate the height of the pole from the length of the shadow. However, one would not want to claim that the length of the shadow somehow causes the height of the flagpole. Causation seems to operate in only one direction. Examples of this kind convinced philosophers of science of the need to reintroduce the notion of causation into the philosophy of science. Since causation is not symmetrical, in the same way that derivation is, the causal direction needs to be included into any account of explanation, or so the post-Positivist story goes.

But more recent work, as already mentioned, has questioned the contemporary hegemony of causal explanations, particularly in the most fundamental discipline of physics (Rivadulla, 2019). But let us assume, for the sake of the present discussion, that there is indeed a strong connection between causation and reduction in the way that Hendry assumes when he writes,

the ontological reductionist thinks that special-science properties are no more than their physical bases because the causal powers they confer are a subset of those conferred by their physical bases; the emergentist sees them as distinct and non-reducible just because the causal powers they confer are not exhausted by those conferred by their physical bases. The additional causal powers are exerted in downward causation (Hendry 2010, p. 185)

Hendry then appeals to the work of C. D. Broad on emergentism and claims that it provides an account of emergence from which a model of downward causation is easily extracted. Writing in the 1920s Broad made a contrast between what he called 'pure mechanism' whereby every material object is made of fundamental particles of one kind of stuff and emergentism where this is not the case. Moreover, according to Broad, one physical law governs the interaction between the particles, and according to pure mechanism, this law determines the behavior of every material object. Hendry's gloss on this point is,

Broad's account of the disagreement between pure mechanism and emergentism is easily formulated within quantum mechanics, in which the motions are governed by Hamiltonian operators determined by the forces acting within a system (Hendry 2010 p. 184).

The notion that such a connection between emergentism and quantum mechanics may be easily formulated also seems rather extravagant. Countless attempts to settle such questions within the philosophy of physics have been highly inconclusive and far from easy. It is by no means clear whether reductionism breaks down in the domain of quantum mechanics.

Hendry also claims that whereas the reductionist posits a resultant Hamiltonian, the emergentist posits a non-resultant Hamiltonian or “configurational Hamiltonian” but unfortunately is unable to identify any such configurational Hamiltonians for the examples which he discusses.

So far, I have largely been summarizing an article which I published in 2012 but which Hendry has yet to respond to (Scerri, 2012). In the same article I suggested that a different alternative, to the existence of emergence, might be to consider the notion that the isomers of any compound, such as one possessing the molecular formula $C_2H_6O_1$, when first formed, might consist of a superposition of its possible isomers. After a very brief period of time the now well accepted process of quantum decoherence might occur so as to collapse the superposition into an actuality featuring one specific isomer. Said in different words, I proposed that at the most fundamental level the initial formation of a molecule really does lack a structure in the sense that it has not yet actualized into a particular structural isomer.

This appeal to the work in the foundations of physics and the question of the collapse of the wave function has been rendered more attractive by the realization that the collapse of the wavefunction can even occur in the absence of observation. All that is required is for there to be an interaction with the environment in which the molecule finds itself in. For example, something as small as a grain of dust is now known to be capable of collapsing the wavefunction (Zurek, 2003). Moreover, research into the foundations of physics has made it possible to compute the decoherence time for any particular molecule, which is typically of the order of femtoseconds. What this amounts to, is the plausible scenario whereby a molecule initially forms as a result of a particular reaction, say the synthesis of $C_2H_6O_1$ and after such a very brief passage of time has elapsed, just one of the two possible structural isomers comes into being.⁵

My proposal for considering the question of the collapse of the wavefunction and quantum decoherence has now been picked up by Seifert and Franklin who have developed a far more detailed account than I could ever have done, as a means to counter any claims as to the occurrence of emergence (Franklin & Seifert, 2023).

8.3 Bonding

The second major theme in the work of Hendry that will be considered is his view of chemical bonding. In previous publications I have suggested that chemical bonding is one of the two big ideas in chemistry, in response to some philosophers of physics who deny any form of philosophical importance to the field of chemistry (Scerri, 2020). Molecular structure and bonding are among the most quintessential topics that have been considered by the new wave of philosophers of chemistry that

⁵ In general, the superposition may involve any number of structural isomers which share the same molecular formula.

began to take shape in the mid 1990s. It is therefore essential that such views be subjected to careful consideration.

The topic of chemical bonding has a long and complicated history, which can be taken to begin with the work of chemist John Dalton at the beginning of the nineteenth century. Dalton revived the atomic theory of the ancient Greek philosophers, some of whom held that matter is not infinitely sub-divisible, but that a limit is reached once one arrives at the atoms, that are the smallest components of each of the elements (Greenaway, 1966).

Dalton proceeded to consider the combination of atoms to form molecules such as water, which he incorrectly believed to consist of one atom of hydrogen combined to one atom of oxygen. The nature of the attraction between these two kinds of atoms was a source of great difficulty for early chemists such as Dalton. In some respects the mystery remains up to the present time, although very accurate calculations on the properties of molecules can now be carried out.

Nevertheless, the question of what chemical bonds actually consist of continues to pose problems and there are many remaining disagreements among professional chemists (Malrieu et al., 2007; Rzepa, 2009).

One of the earliest views that was contemplated was that chemical bonds are physical links between the constituent atoms. These physical connections were thought to be stick like linkages or perhaps in the form of mechanical springs. Stated otherwise, bonding was originally viewed in a naïvely realistic sense of physical entities which were as substantial as the atoms that they were thought to connect together.⁶

In the early part of the twentieth century great advances were made, resulting in the classification of chemical bonds into the categories of ionic and covalent bonding. Ionic bonding was postulated first to consist of an attraction between charged ions, resulting from the complete transfer of electrons from metal atoms to atoms of non-metals. The ions formed in this way were considered to attract each other and to form three-dimensional crystal lattices, such as in the classic example of sodium chloride (Kossell, 1916). Soon afterwards an alternative form of bonding was proposed by G.N. Lewis, in order to explain the existence of non-polar compounds, in which oppositely charged ions did not play any role (Lewis, 1916). This other major form of bonding was called covalent bonding in order to reflect the notion that constituent atoms were sharing electrons rather than transferring them. Examples include such molecules as diatomic gases such as H₂, O₂ and so on. For about 100 years schoolchildren have been learning the basic distinction between these two kinds of chemical bonds right from the beginning of their chemistry courses.

As in the case of most elementary ideas in science, this simple picture must be qualified as instruction in the subject is taken to more advanced levels. For example, one must appreciate the fact that the two forms of bonding are but extremes on a single continuous spectrum. It is more helpful to think of the two forms of bonding

⁶To the extent that atoms were regarded as real physical entities, a view that was by no means universal among chemists such as Mendeleev and many others, especially in the eighteenth and nineteenth centuries.

as being cases of approximately equal sharing of electrons in the case of covalent bonding, as compared with very unequal sharing of electrons in the ionic case.⁷ Any philosophical analysis which is predicated on the characteristic difference between ionic and covalent bonding is therefore problematical from the outset, a feature which I believe has occurred in some of the recent discussion in the philosophy of chemistry community, as I will attempt to explain.

8.4 Hendry's Contrast Between the Energetic and the Structural View of Bonding

In a further series of articles Robin Hendry has written about what he considers to be opposing views concerning the nature of chemical bonding. Hendry's 'structural conception' of chemical bonding consists of the claim that a covalent bond is a directional, sub-molecular relationship between individual atomic centers, that is responsible for holding the atoms together. However, even in classical chemistry covalent bonding is not invariably directional and it is not necessarily sub-molecular, although I will delay a fuller discussion of these points for the moment.

It is well-known that the distinction between ionic and covalent bonding is something of an over-simplification. The modern study of chemical bonding frequently involves the application of the Schrödinger equation for the physical system in question and in so doing one does not pause to specify whether the bonding might be ionic or covalent. Give this state of affairs there would seem little point in attempting to specify the quintessential nature of just covalent bonding.

Further aspects of the Hendry's structural conception consist in the notion that ionic bonds are omnidirectional electrostatic interactions between positively and negatively charged ions while covalent bonds are regions of electron density that bind atoms together along particular trajectory.

The second sentence would seem to imply that ionic bonds do not involve regions of electron density, which is surely not what Hendry means to say. As to the question of directionality, this characterization would seem to omit an entire class of covalently bonded compounds such as diamond or graphite in which bonding is multi-directional just as in classic cases of ionic bonding.

Another philosopher of chemistry, Weisberg, drawing on Hendry, writes that,

Second, this [structural] conception says that bonding is a *sub-molecular* phenomenon, confined to regions between the atoms. This eliminates the possibility that bonds are a molecule-wide phenomenon (Weisberg, 2008, 935).

If this is intended as a further characteristic of just covalent bonding it is simply incorrect, since ionic bonding also occurs between atoms, or more correctly their ions. I am also puzzled by the apparent desire to exclude the possibility that bonds, or bonding, might be a molecule-wide phenomenon. Counter examples are easy to

⁷This point was already emphasized by G.N. Lewis almost exactly 100 years ago.

find. In addition to diamond and graphite, which are generally described as displaying giant covalent bonding, modern chemistry has revealed the frequent occurrence of delocalized bonding to occur in cases such as metals, conducting polymers, benzene and many other conjugated hydrocarbons. Moreover, delocalization of electrons is known to occur in many inorganic species such as oxyanions including the carbonate and sulfate ions. Bonding is indeed a molecule-wide phenomenon and delocalization is not confined to covalent compounds.

Thirdly, Robin Hendry⁸ believes that an article published by the late Gerome Berson provides support for own his view that bonds really exist between any particular two atoms in any molecule.⁹ In this article Berson reports on some unusual molecules which seem to support the notion that the energetic view of bonding is problematical. It should be emphasized that this conclusion was not in fact drawn by the author Berson but only by Hendry. The molecules in question are one labeled 9T which Berson compares with molecule 11 as shown in Fig. 8.1.

It appears that the more stable of the two molecules, 9T, possesses fewer bonds, as understood in the classical sense of the sharing of two electrons between any two given atoms. For Hendry this seems to indicate a violation of the equivalence between the extent of bonding and achieving the most stable energy. Molecule 9T appears to be more stable even though it has fewer bonds than molecule 11. Hendry's conclusion is that the energetic view provides an incomplete picture and that the structural view therefore appears to be superior in this instance.

I would like to propose looking at this issue from a different perspective. The fact that the molecule with fewer bonds is the more stable of the two, serves to illustrate that the naïve picture of 2 electrons to each bond between specific atoms might be where the problem lies. Far from supporting Hendry's position the molecule that

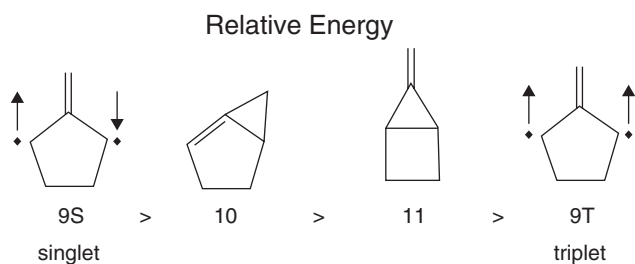


Fig. 8.1 Relative energies of singlet and triplet 2-methylenecyclopentane-1,3-diyl and their precursors (Berson, 2008, 951). (Reproduced with permission)

⁸Private E-mail correspondence with Robin Hendry.

⁹The article was based on a lecture given by Gerome Berson at the same session of the Philosophy of Science Association at which Hendry and Weisberg spoke in 2007. The only philosophical comment that Berson makes in his article is that, "Chemists therefore seek to enlist philosophers in sharpening the very definition of a bond." (p.947)

Berson has described, exposes the superficial nature of regarding bonds as specific inter-nuclear entities.

What this unusual molecule 9T shows, if anything, is that there appears to be a greater degree of ‘bonding’ despite the fact that there are fewer specific bonds in the naïve sense of the organic chemist. The degree of bonding in general thus remains correlated to the degree of energy minimization.¹⁰

Moreover, Berson’s analysis supports the view that the stability of molecules is the more important factor in considering the interconversion of molecules, regardless of precisely how many bonds are present in the classical sense of pairs of shared electrons. Or to cite Berson,

The bond concept allows us to understand much of chemistry, but far from all of it (Berson, 2008, 954).

Finally, I turn to an issue that represents perhaps the greatest threat to Robin Hendry’s view concerning the importance of the structural view, and his belief that bonds are ‘real’ in some unspecified way. In order to discuss this aspect, one must consider the quantum mechanical account of the covalent bond.

8.5 Quantum Mechanical Account of the Covalent Bond

Soon after Schrödinger published his wave equation for the hydrogen atom, two young post-doctoral fellows, Heitler and London, succeed in calculating the energy of the simplest molecule, H₂, and in showing that it was stable. In order to do so they drew on the fact that electrons acting through their wave nature would interact via constructive and destructive interference. The result of constructive interference is generally believed to be a build-up of electron density between the nuclei on adjacent hydrogen atoms, such that the two electrons that are shared in the covalent bond can be regarded as a form of ‘glue’ that causes the two positive nuclei to be attracted to each other. One apparent advantage of this interpretation is that it accords very well with the previous view of G.N. Lewis, namely that a covalent

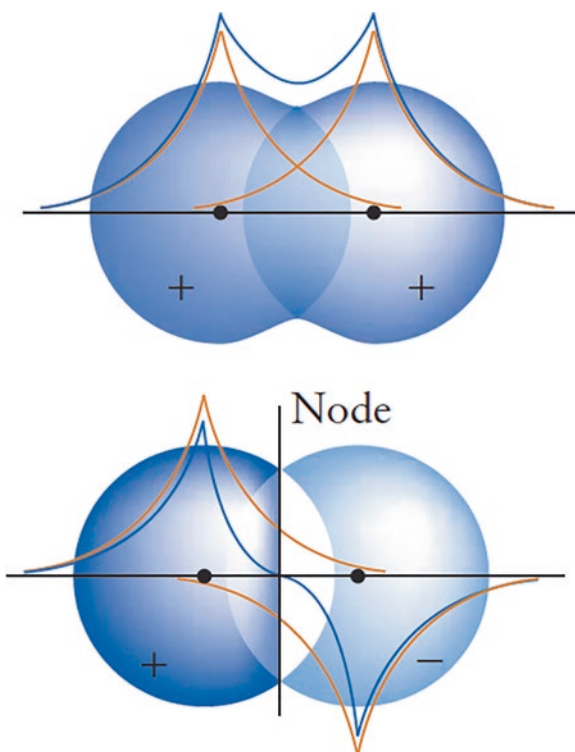
¹⁰In another figure, labeled 6, Berson connects structure 11 with structure 9S (a singlet species but having two unpaired electrons) over a transition state. Berson also connects 10 with 9S over another transition state. The author does not connect the 9T structure (a triplet species having two unpaired electrons) to any of the others because the triplet and singlets are of different symmetry and do not therefore couple or connect. It should also be noted that although structure 9T has fewer bonds than 11 or 10, (i) 9T has less internal strain energy within its ring than 11 (and maybe 10) and (ii) 9T is a triplet while the others are singlets. The importance of the latter statement can be appreciated by noting the 9S (which has fewer bonds just like 9T does) is actually higher in energy than 11 or 10. So, within the singlet world, the systems with more bonds (11 and 10) do indeed have lower energy than the system with fewer bonds (9S). The relative orderings of various structures depend on intrinsic bond strengths but also on strain energies as well as the energy difference between unpaired electrons in singlet or triplet couplings. I am grateful to Professor J. Simons for discussion on these issues.

Fig. 8.2 The conventional textbook explanation of bonding correctly begins by considering electron waves on adjacent atoms which combine together constructively and also destructively



The electron waves on each atom combine constructively as well as destructively.

Fig. 8.3 The top part of the image depicts constructive interference between waves on adjacent atoms leading to an increase in electron density between the nuclei. The lower part of the diagram depicts out-of-phase interaction leading to the depletion of electron density between the nuclei (Permission requested)

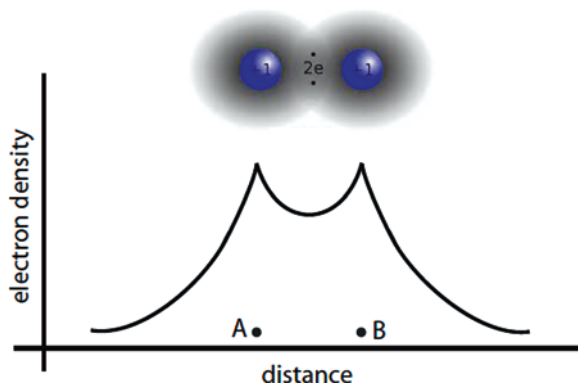


bond consists of a pair of electrons located mid-way between the two hydrogen atoms in the H_2 molecule.

The traditional interpretation of the quantum mechanical theory of chemical bonding arises from treating the electron as a wave and considering the interaction of the waves between two adjacent hydrogen atoms as shown in Fig. 8.2.

When any kind of waves combine together, they give rise to constructive as well as destructive interference. The former case results in a build-up of electrons between the nuclei. At the same time the destructive interference results in the depletion of electron density between the adjacent atoms as can be seen in Fig. 8.3.

Fig. 8.4 Constructive interference of electron waves as depicted in the upper part of Fig. 8.3 and the correspondence with the classical notion of a shared pair of electrons situated between adjacent atoms



The familiar textbook explanation of chemical bonding focuses primarily on the constructive interference contribution which serves to recover very much the same kind of picture of bonding as was first proposed by G.N. Lewis, namely that a covalent bond consists of a pair of electrons that are shared between adjacent atoms as illustrated in Fig. 8.4.

This conception of covalent bonding is somewhat erroneous since it ignores the contribution arising from destructive interference of the electron waves. Moreover, it is essentially an electrostatic view which ignores any contributions from the kinetic energy of the electrons. Whereas the calculations carried out by the likes of Heitler and London included kinetic energy terms in the Hamiltonian of the molecule, the simplified picture that we are discussing here would seem to be focusing exclusively on the potential energy contribution which is essentially static. The very notion of an electron glue situated in a particular location between the nuclei reinforces the notion of a static rather than dynamical view.

Fortunately, there is a long-standing line of argumentation among theoretical chemists that challenges this naïve notion. Beginning in the 1930s Hellman pioneered the view that covalent bonding was dominated by contribution of the kinetic energy of the electrons rather than their potential energy (Hellmann, 1937). For many years this view was ignored by most theoretical chemists until it was reformulated in a more rigorous fashion by the theoretical chemist Klaus Ruedenberg (Ruedenberg, 1962; Ruedenberg & Schmidt, 2007).

In order to illustrate the main ideas in the Hellman-Ruedenberg approach I now turn to an even simpler molecule than H_2 , namely the H_2^+ ion in which just a single electron is shared by the two adjacent hydrogen nuclei.¹¹ The Hamiltonian for this system is shown in Fig. 8.5.

In addition to calculating the total energy of the H_2^+ molecule-ion, it is possible to calculate the separate contributions due to kinetic and potential energy arising from the bonding and anti-bonding contributions due to constructive and destructive

¹¹The fact that this molecule-ion contains chemical bonding immediately belies the simple notion due to Lewis that a covalent bond consists of a *pair* of shared electrons.

The hydrogen molecule-ion

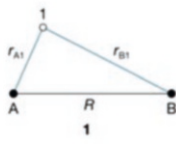
$$H = -\frac{\hbar}{2m_e} \nabla_1^2 + V \quad V = -\frac{e^2}{4\pi\epsilon_0} \left(\frac{1}{r_{A1}} + \frac{1}{r_{B1}} + \frac{1}{R} \right)$$


Fig. 8.5 The Hamiltonian operator for the H_2^+ molecule-ion, in which V represents the potential energy which is made up of three terms

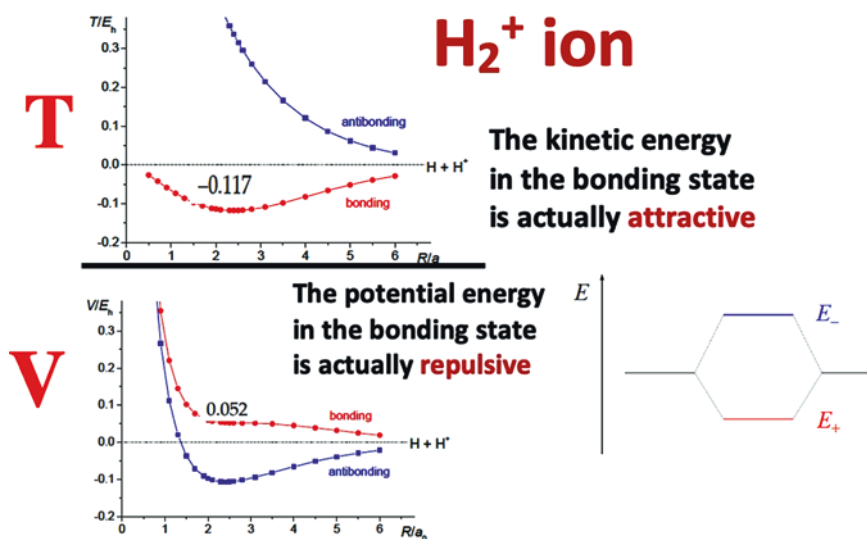


Fig. 8.6 Graphs of kinetic energy (T) and potential energy (V) as a function of internuclear separation. The attractive force is only present in the kinetic energy contribution to the total energy. (Diagram modified from Bacskay et al. (2010) and private correspondence with G. Bacskay)

interference respectively. The results of such calculations are displayed in Fig. 8.6 below. These graphs show very clearly that the attraction between the two hydrogen nuclei is due to the negative kinetic contribution and not to the contribution from the potential energy, which is in fact a positive and hence a repulsive term. The graphs also show that if the force responsible for bonding in this molecule-ion was due to potential energy alone, it would not lead to any bonding whatsoever, and there would be no means to overcome the repulsive force between the two positively charged hydrogen nuclei.

Given these facts it becomes difficult to maintain the classical view that a covalent bond consists of the sharing of electrons which are located between two adjacent atoms. More importantly for the main theme of the present article, it presents a major challenge for what Hendry has termed the structural view of chemical bonding which aims to recover directional bonds between particular atoms in a molecule and the notion that such bonds are somehow ‘real’. A more correct view according to the kinetic energy school of thought is to suppose that chemical bonding, rather than specific bonds, as such, is the result of electrons that are shared *by* nuclei but that do not necessarily lie between them. According to this view electrons are being shared by two or more atoms but not between these atoms.

Here is the way that one author expresses the alternative view of bonding,

The amount of electron density transferred to the bonding region is greatly overstated, sometimes implying that a pair of electrons is shared in the space between two nuclei rather than by two nuclei (Rioux, 2003).

8.6 Are Bonds Real?

A major pre-occupation for Hendry, among several philosophers of chemistry, has been the question of whether bonds are ‘real’ (Hendry, 2008; Weisberg, 2008; Seifert, 2022). For example, Hendry has attempted to refute the view of authors like Coulson who claimed that,

a bond ‘does not exist: no-one has ever seen it, no-one ever can. It is a figment of our own imagination (Coulson, 1955)

by appeal to Bader’s theory of atoms in molecules in which bond paths, rather than bonds, are a central feature of the theory. But as Hendry readily concedes, Bader’s view raises several conceptual problems, among them being the fact that it sometimes shows the presence of bond paths where they clearly cannot exist and in other instances represents a repulsive interaction as a bond path.

As Bader puts it, “The recovery of a chemical structure in terms of a property of the system’s charge density is a most remarkable and important result” (1990, 33). But the correspondence between bond path and chemical bond is not perfect. The main problems concern repulsive (rather than attractive) interactions between neighbouring atoms in a molecule. Bader’s algorithm finds bond paths corresponding to these repulsive interactions, even though chemists would not normally regard the mutually repelling pairs of atoms as bonded to each other (Hendry, 2018, 113).

To conclude this section, I believe that the debate concerning the reality of bonds and the supposed opposition between the structural and energetic views are both vacuous. The alleged debate between the structural and the energetic view is essentially a return to the debate among chemists over the superiority of the valence bond or molecular orbital theories. Whereas bonds are regarded as real in the valence bond approach, molecular orbital theory assumes the presence of delocalized bonding but not specific bonds. The two approaches were shown to be completely

equivalent to each other by Slater and Van Vleck as early as the 1930s. Consequently there is no longer any debate over this question. Some philosophers of chemistry including Hendry and Weisberg are merely attempting to revive the same debate by asking the metaphysical question of whether bonds are 'real', but this does not alter the central issue.

As Paul Needham has written,

Construing the status of the chemical bond as an issue of existence, is perhaps an unfortunate formulation. What exists are entities such as molecules, atoms and electrons, whereas bonding is something they do. The question is How? (Needham, 2014, 11).

8.7 Conclusions

Hendry promotes the continuity of the concept of bonding. One may well agree with this notion of continuity in scientific concepts, as I have argued in a previous publication (Scerri, 2016). However, there is no denying that talk of bonds has now morphed into talk of 'bonding' in the quantum mechanical account. Bonding is now discussed in energetic grounds rather than via a realistic belief in entities that connect atoms together.

The notion that there are in fact two views of bonding is a relic of a debate that took place in the 1950s. The energetic view does indeed prevail over the structural view, if one must speak in these terms. Said otherwise, Slater and Van Vleck showed some 90 years ago that the valence bond and molecular orbital theories are completely equivalent mathematically (Slater, 1932; Van Vleck & Sherman, 1935).¹² Of course organic chemists may continue to regard bonds as pairs of electrons and may also think of molecular structure as being irreducible to quantum mechanics for the sake of expediency, but this does not sanction the ontological claim made by Hendry, to the effect that the structural and energetic views are still competing among each other as to which of them is the more correct description of chemical bonding.

References

- Bacskay, G. B., Eek, W., & Nordholm, S. (2010). Is covalent bonding a one-electron phenomenon? Analysis of a simple potential model of molecular structure. *The Chemical Educator*, 15, 42–54.
- Berson, G. (2008). Molecules with very weak bonds: The edge of covalency. *Philosophy of Science*, 75(5), 947–957.
- Coulson, C. A. (1955). The contributions of wave mechanics to chemistry. *Journal of the Chemical Society*, 2069–2084. <https://doi.org/10.1039/JR9550002069>

¹²Also see Galbraith et al. (2021) for a recent statement on the equivalence of the two theories and the manner in which they are frequently misrepresented in chemistry textbooks.

- Franklin, A., & Seifert, V. A. (2023). The problem of molecular structure just is the measurement problem. *British Journal for the Philosophy of Science*, 75. <https://doi.org/10.1086/715148>
- Galbraith, J. M., Shaik, S., Danovich, D., Braïda, B., Wei, W., Hiberty, P., Cooper, D. L., Karadakov, P. B., & Dunning, T. H., Jr. (2021). Valence bond and molecular orbital: Two powerful theories that nicely complement one another. *Journal of Chemical Education*, 98(12), 3617–3620.
- Gambini, R., & Pullin, J. (2020). *Loop quantum gravity for everyone*. World Scientific Press.
- Greenaway, F. (1966). *John Dalton and the Atom*. Cornell University Press.
- Hellmann, H. (1937). *Quantenchemie*. Deuticke, Leipzig and Wien.
- Hendry, R. (2008). Two concepts of chemical bond. *Philosophy of Science*, 75, 909–920.
- Hendry, R. F. (2010). Ontological reduction and molecular structure. *Studies in History and Philosophy of Modern Physics*, 41, 183–191. <https://doi.org/10.1016/j.shpsb.2010.03.005>
- Hendry, R. (2018). Scientific realism and the history of chemistry. *Spontaneous Generations: A Journal for the History and Philosophy of Science*, 9(1), 108–117.
- Jammer, M. (1966). *The conceptual development of quantum mechanics*. McGraw-Hill.
- Kossel, W. (1916). Molecule formation as a question of atomic structure. *Annalen der Physik*, 49, 229–362.
- Lange, M. (2013). What makes a scientific explanation distinctively mathematical? *British Journal for the Philosophy of Science*, 64, 485–511.
- Lewis, G. N. (1916). The atom and the molecule. *Journal of the American Chemical Society*, 38(1916), 762–786.
- Malrieu, J.-P., Guihéry, N., Jiménez Calzado, C., & Angeli, C. (2007). Bond electron pair: Its relevance and analysis from the quantum chemistry point of view. *Journal of Computational Chemistry*, 28, 35–50.
- Matta, C. (2023). Causal mapping: Some observations and questions by a chemist, A lecture delivered at the 2023 conference of Congress on Logic, Methodology and Philosophy of Science and Technology, Buenos Aires, Argentina (article in preparation).
- Mehra, J., & Rechenberg, H. (1987). *The historical development of quantum theory: Erwin Schrodinger and the rise of wave mechanics: The creation of wave mechanics early response and applications 1925–1926*. Springer.
- Needham, P. (2014). The source of chemical bonding. *Studies in History and Philosophy of Science*, 45, 1–13.
- Norton, J. D. (2003). Causation as folk science. *Philosophers' Imprint*, 3(4), 1–22.
- Primas, H. (1983). *Chemistry, quantum mechanics and reductionism perspectives in theoretical chemistry*. Springer.
- Rioux, F. (2003). The covalent bond examined using the virial theorem. *The Chemical Educator*, 8, 10–12.
- Rivadulla, A. (2019). Causal explanations: Are they possible in physics? Causal explanations: Are they possible in physics? In M. R. Matthews (Ed.), *Mario Bunge: A centenary festschrift*. Springer.
- Ruedenberg, K. (1962). The nature of the chemical bond. *Reviews of Modern Physics*, 34, 326.
- Ruedenberg, K., & Schmidt, M. W. (2007). Why does electron sharing lead to covalent bonding? A variational analysis. *Journal of Computational Chemistry*, 28, 391–410.
- Rzepa, H. (2009). The importance of being bonded. *Nature Chemistry*, 1, 510–512.
- Scerri, E. R. (2012). Top-down causation regarding the chemistry – Physics interface – A skeptical view. *Interface Focus*, Royal Society Publications, 2, 20–25.
- Scerri, E. R. (2016). *A tale of seven scientists and a new philosophy of science*. Oxford University Press.
- Scerri, E. R. (2020). *The periodic table, its story and its significance*. Oxford University Press.
- Scerri, E. R. (2021). Causation, electronic configurations and the periodic table. *Synthese*, 198, 9709–9720.
- Seifert, V. (2022). The chemical bond is a real pattern. *Philosophy of Science*, 1–47. <https://doi.org/10.1017/psa.2022.17>. Published online, 22nd April.
- Slater, J. C. (1932). Note on molecular structure. *Physics Review*, 41, 255–257.

- Van Vleck, J. H., & Sherman, A. (1935). The quantum theory of valence. *Reviews of Modern Physics*, 7, 167–228.
- Weisberg, M. (2008). Challenges to the structural conception of chemical bonding. *Philosophy in Science*, 75, 932–946.
- Woolley, G. (1976). Quantum theory and molecular structure. *Advances in Physics*, 25, 27–52.
- Zurek, W. H. (2003). Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75, 715–775.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 9

Fundamental Physics and (New-) Mechanistic Ontologies



João L. Cordovil

Abstract According to Kuhlmann & Glennan, fundamental physics and New Mechanicism do “not fit well together” (Kuhlmann and Glennan, *Euro J Phil Sci* 4:338, 2014). For two main reasons: (1) Quantum mechanics (QM) challenges the hypothesis that there are objects with definite properties that are related by local causal interactions; (2) since mechanisms are composed of lower-level mechanisms, then if in fundamental physics the existence of mechanisms can be questioned, and if macroscopic mechanisms supervene on fundamental physics entities and processes, then fundamental physics can even undermine mechanistic ontology and its explanatory ambition.

In their paper, Kuhlmann & Glennan tried to argue that the problem of the compatibilisation between fundamental physics and New Mechanicism can be partially addressed since, on the one hand, the quantum decoherence hypothesis allows to defend that the universal validity of quantum mechanics does not undermine New Mechanicism ontological and explanatory claims as they occur within in classical domains. And on the other hand, it is possible to offer a non-classical mechanistic explanation of certain kinds of quantum phenomena.

This paper aims to argue that there has always been a problematic relationship between mechanical philosophy and fundamental physics throughout the history of physics. Therefore, in part, the challenges posed by QM to mechanicism are not new; nevertheless, mechanicism prevailed throughout the history of physics. On the other hand, I also aim to argue that although fundamental physics may not be compatible with New Mechanicism, that should not imply a rejection of mechanistic ontology for reasons other than the quantum decoherence hypothesis.

Keywords Action at a distance · New mechanism · Mechanical philosophy · Ontological emergence · Entanglement

J. L. Cordovil (✉)

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_9

179

9.1 Introduction

In the last years, some attention has been given to what has been called “New Mechanicism” or “New Mechanical Philosophy” (Craver & Tabery, 2019; Glennan, 2017). One of the primary motivations for New Mechanicism is that in opposition to some tradition in the philosophy of science, theories and universal laws are not the basis for explanations in science – for instance. The discussion occurs within a naturalised approach to metaphysics; however, as noticed by Kuhlmann and Glennan (2014), despite increasing interest among science philosophers, physics has never been the focus of more significant debates concerning the New Mechanicism. That may come as a surprise, not only because of the central role that physics always has taken in the philosophy of science but considering the long tradition that mechanisms and Mechanical Philosophy have in physics. The simplest explanation for this absentness is that fundamental physics not only may not be compatible with the mechanistic approach, but even more dramatically, fundamental physics may undermine the mechanistic program in science. According to Kuhlmann and Glennan (2014):

There is *prima facie* reason to be concerned that the two pictures do not fit well together. The neo-mechanists suppose that mechanisms are composed of objects with definite properties, where these objects are connected via local causal interactions. Quantum mechanics (QM) calls into question whether there are such things as objects with definite properties and whether causal relations can be understood in terms of local interactions between such objects. Moreover, mechanisms are hierarchical in the sense that the parts of mechanisms may themselves be complex objects composed of subparts which are components of lower-level mechanisms. It seems then that even complex macroscopic mechanisms must supervene on a set of “objects” that behave non-classically. This dependence upon a non-classical micro-level might seem to infect the ontological and even explanatory claims of the New Mechanists.

Thus, considering the above description, there are four suppositions are still made by neo-mechanists about mechanisms: (1) that there are objects with definite properties; (2) these objects are connected to each other through causal interactions; (3) relations can be understood in terms of local interactions between objects; (4) there is a hierarchical structure between mechanisms.

Taking these assumptions, the problem, the clash between neo-mechanists and fundamental physics would be the “fact” that fundamental physics challenges all the four suppositions stated above. If this is the case, then, as stated by Kuhlmann & Glennan, it seems that not only we can be sceptical about the existence of mechanisms in fundamental physics, but since macroscopic mechanisms must supervene on fundamental physics entities and processes, and if these do not fit into the mechanistic picture, then fundamental physics can even undermine mechanistic ontology and ambition. Kuhlmann clearly restates that:

The primary concern is not how the notion of mechanisms can be captured and how exactly mechanistic explanations work. Rather, the main question is to what extent physics, in particular fundamental physics, deals with mechanisms in the first place. A second question concerns whether the character of the physical processes that underlie all natural and social phenomena may even endanger the tenability of mechanistic reasoning in the special sciences. Kuhlmann (2018)

This raises at least the following questions: (1) are the problems that fundamental physics places on the New Mechanical Philosophy any new in the history of physics? (2) Even if it is the case that fundamental physics is incompatible with mechanisms, why does it place a problem with the suitability of the New Mechanistic approach on special sciences?

9.2 Traditional Mechanical Philosophy in Physics

9.2.1 *Descartes*

Mechanical Philosophy has a long tradition in Physics. It can even be said that his foundational philosophy, both in the ontological, epistemic, and methodological sense. Beginning with Descartes. As Garber notes:

“Descartes saw the physical world and its contents as a collection of machines. At the end of his *Principia Philosophiae*, Descartes tell the reader that “I have described this earth and indeed the whole visible universe as if it were a machine: I have considered only the various shapes and movements of its parts” (Pr IV, 188). Later in the *Principia* he writes:

“I do not recognize any difference between artifacts and natural bodies except that the operations of artifacts are for the most part performed by mechanisms which are large enough to be easily perceivable by the senses—as indeed must be the case if they are to be capable of being manufactured by human beings. The effects produced in nature, by contrast, almost always depend on structures which are so minute that they completely elude our senses. (Pr IV, 203)”

Similarly, Descartes suggests to an unknown correspondent, seeking to clarify his position that “all the causes of motion in material things are the same as in artificial machines.”” (Garber, 2013: 16)

According to Descartes, physical objects or bodies are *Res Extensa*. That is, physical objects are similar to three-dimensional shape-geometrical objects, with length, breadth, and depth. Therefore, there is a clear and radical cut between mental and physical properties, between subjects and objects.

Since all physical objects are extended entities, there are no atoms; all physical objects must be divisible, no matter their size. However, for the same token, there cannot exist extension without physical objects; that is, there are no atoms nor empty space. Therefore, all space is completely “filled” by bodies and material objects; Space is a *plenum*. Space is the composition of physical objects of different sizes. Some are so tiny, that has an indefinite extension – *corpuscula*.

Consequently, every physical object is in direct contact with its surrounding bodies and, indirectly, with all other objects, *via* the imbricate of corpuscula that composes the *plenum*. Hence, all physical objects are deeply intertwined, and any motion of one physical object is necessarily communicated to the others. It is the image of the natural world in Descartes: a colossal clock with nothing more to consider than shape and motion. A machine that underlies and grounds all physical phenomena. A machine in which each part, or different sub-machines, works together in complete harmony and crosses all compositional levels. The natural

world (physical, biological, etc.) is the arrangement of interrelated geometrical parts (like cogs in a clock). The scientific task is to analytically decompose any phenomenon or body into its components since all phenomena is fully explained by the causal and local interaction of physical objects, according to the laws of physics. All natural phenomena can and should be explained by the motion and collision of particles of matter and its composition alone. This is the epitome, the zenith, of the mechanistic philosophy in physics.

However, on the one hand, motion is a necessary condition to explain all phenomena, but why the universe, the “clock”, and the “machine” started to move in the first place is not explainable in mechanistic terms. Even in Descartes, Mechanical Philosophy is not self-sufficient.

9.2.2 *Newton*

The decisive challenge to the mechanistic approach in physics came with Newton. As put by Cohen (1999:57):

“Newton was still “a mechanical philosopher in some sense,” but not any longer in the strict sense in which that designation was usually understood. Whereas a strict mechanical philosopher sought the explanation of all phenomena in terms of what Boyle once called those “two grand and most catholic principles of bodies, matter and motion”, Newton came to believe that “the ultimate agent of nature would be . . . a force acting between particles rather than a moving particle itself”.

On the one hand, likewise traditional mechanical philosophers, such as Descartes or Boyle, Newton regarded the universe as a machine ruled by universally applicable axioms. Axioms that could be discovered by scientific analysis and that would ground the explanation of physical phenomena. In fact, that would be the fundamental role of science or experimental philosophy. Like traditional mechanical philosophers, the basic ontology is constituted by movable material bodies with extension (shape and size), hardness and impenetrability, as explained by Newton in Rule III of the Rules of Reasoning in Natural Philosophy of the *Principia*.

On the same path, in the introduction of the *Principia*, there is a moment where Newton seems to be hopeful concerning the possibilities of Mechanical Philosophy:

For in book 3, by means of propositions demonstrated mathematically in books 1 and 2, we derive from celestial phenomena the gravitational forces by which bodies tend toward the sun and toward the individual planets. Then the motions of the planets, the comets, the moon, and the sea are deduced from these forces by propositions that are also mathematical. If only we could derive the other phenomena of nature from mechanical principles by the same kind of reasoning! (Newton (1999): 382)

However, it is unclear how Newton’s characterisation of forces can be integrated into a mechanistic ontological view. As Janiak (2021) stated:

His second law indicates that a body moving rectilinearly will continue to do so unless a force is impressed on it. This is not equivalent to claiming that a body moving rectilinearly will continue to do so unless another body impacts upon it. A *vis impressa*—an impressed force—in Newton’s system is not the same as a body, nor even a quality of a body, as we have seen; but what is more, some impressed forces need not involve contact between bodies at all.

The main concern is gravity – of course. Gravity is both a kind of central force and an impressed force. Thus, a body moving in a straight line will be instantly deviated by a gravitational force that was originated by another material body placed far away without any intervention (collision) of another body(ies). Therefore, as Janiak (2014: XXVII) recalls:

These elements of the *Principia* make conceptual room for a causal interaction between two bodies separated by a vast distance, one enabled by Newton’s concept of an impressed force. Aspects of this idea became known in philosophical circles as the problem of action at a distance.

Hence, one of the significant successes of the *Principia* – gravitational force – is not explained by any underlying mechanism. Of course, the same could be said about hardness, for instance—or inertial force. However, the gravitational force is more striking since, not only because it is the only force in the *Principia* with a specific law but because it is a force that acts at a distance. That is, a mass A instantly changes the state (of motion) of a mass B, and simultaneously, its state is changed by a mass B, without direct contact (collision, for instance) or by the transmission of force (throughout other masses or a *medium*) between them.

Why is this pose a problem to Mechanical Philosophy, and what is its relationship with contemporary fundamental physics?

The term mechanical in the context of Mechanical Philosophy meant (see, for instance, McGuire (1972)) many different things throughout the history of philosophy. Nevertheless, contact action was, in general, broadly accepted as a necessary (although not sufficient) condition for a mechanical explanation. Thus, if the only thing that is clear for most mechanists is that contact action is necessary, then since gravitational force in Newton is an “action at a distance” and therefore is not compatible with the mechanistic philosophy, then, as Leibniz would put it “*Principia* renders gravitation a “perpetual miracle” because it does not specify the physical mechanism underlying it” Janiak (2021). Alternatively, as also Andrew Janiak (2008: 53) puts “If Newton contends that gravity exists, he must admit that material bodies act on one another at a distance, thereby violating a crucial norm of the mechanical philosophy (in all its guises).”

So, with Newton’s axioms of movement and with the description of the world displayed in books I and II of the *Principia*, on the one hand, we find a realisation of the mechanical philosophers’ ideal: all phenomena seem to be explainable uniquely by the knowledge of the position and the *momenta* of material bodies, and its laws of *momentum*’ determination, transmission, and conservation (The Three Axioms of movement). All phenomena seem to be explainable by the application of this simple ontology. However, on the other hand, Newton’s account of gravitational force encompasses the instant action of a force on a distant matter without any intermediation, something that Mechanical Philosophy cannot follow.

9.3 Contemporary Fundamental Physics and (New) Mechanical Philosophy

9.3.1 Entanglement

One of the most challenging QM's features to traditional Mechanical Philosophy is entanglement. The term "entanglement" was originally coined by Schrödinger (1935) and referred to a special case "where two (or more) particles exist in an eigenstate of a certain observable, such as angular momentum, but neither particle is in an individual eigenstate of that observable" (Huang, 2007: 62).

In the literature, it is possible to find several examples of quantum entanglement. Probably, the most well-known is the one from Bohm (1951: 611–622). Giving a very simplified version of that example (Cordovil, 2015), consider a system in a spin-zero state that decays into a pair of two particles, namely, two electrons—electron A and electron B—that head off in directed opposite directions. Since they are electrons, they have a half-value of spin. Using the standard convention, the spin state is either "up" (+1/2) or "down" (−1/2).

In this case—and only considering the spin factor—there are two independent spin wave functions α e β , representing respectively the state "up" and the state "down". In this state the total spin is definite—singlet state—but the individual spins (of the electrons 1 and 2) are not defined. All we can know, by the conservation of angular *momentum*, is that if one is in the spin-state "up", the other will be in the spin-state "down", and vice-versa. More specifically, if electron A acquires (or shows) the value spin "up", electron B will acquire (or show), through measurement, the value spin "down" immediately and no matter the distance (and vice versa). That is, the spin state of electron A is dependent on the spin state of electron B, and vice-versa. This is the reason why it is said that particle A and particle B are entangled. The mystery is: how can one account for something that was at one point indefinite regarding its spin (or whatever the property under investigation) and that suddenly becomes definite even though no physical interaction (direct or indirectly) with the other subsystem occurred? How, instantly, can the interaction via measurement with one particle immediately alter, at a distance, the state of the other particle?

In a way, the challenges posed to Mechanical Philosophy by the Newtonian description of the force of gravity, until the advent of General Relativity, seem to be not much different to those posed by quantum entanglement. In both cases there is an instantly interaction between two spatially separated physical objects, without the mediation of other objects or entities. Furthermore, in both cases the change of the state of one object instantly changes the state of the other object, no matter the distance. Thus, if one of the critical elements of Mechanical Philosophy is that the world is composed of local causal interactions, then, since Newton, that does not seem to be compatible with fundamental physics. Consequently, maybe apart from the brief period between the appearance of general relativity and the formulation of quantum mechanics, Mechanical Philosophy always had a problematic relationship with fundamental physics. Then, the reason why physics has not been part of the

contemporary discussions on New Mechanicism should not be because there is a clash between fundamental physics and New Mechanicism, as was defended by Kuhlmann & Glennan. First, because that clash has always been present in the history of physics. Secondly, because general relativity is also fundamental physics and the abovementioned clash between physics and New Mechanicism may not exist. Such analysis was not made.

It could be argued that the conflict between fundamental physics and Mechanicism is more acute in QM than with Newtonian gravitation. Although the force of gravity has been a mystery for the Mechanical Philosophy advocates, there was always the expectation that gravity would eventually be explained mechanistically. For instance, Leibniz, Boshokovic or even Kant strongly reacted against the non-mechanical character of gravitational force and tried to offer alternatives. Due to this well-known discomfort, one could defend that the non-local feature of Newtonian has not had the same status that entanglement does since the physicists and philosophers of physics alike accept the latter. However, that is not the case. For instance, some interpretations or reformulations of quantum mechanics try to incorporate entanglement into a mechanistic framework. This could happen, for example, by defending a local interpretation of quantum mechanics.

The de Broglie-Bohm theory, also known as the pilot-wave theory, proposes that particles have definite positions and trajectories, where the wave function serves as a guide or “pilot wave” that determines how particles move. The theory is fully deterministic and does not require any non-local effects. Another example would be the Many-Worlds Interpretation, where the wave function describes a “multiverse” in which every possible measurement outcome occurs in a separate parallel universe. Alternatively, any other interpretation that incorporates the idea of hidden variables, where the wave function is not a complete description of the system but instead reflects our lack of knowledge (for instance Araújo et al., 2009; Lopez, 2016). Or a relational kind of interpretation of QM along the lines of Esfeld’s Thin-Objects Moderated Ontic Structural Realism (Esfeld et al. (2015)).

So, in fact, as in the case of gravitational force, even though our standard fundamental physics may not be compatible with (New) Mechanical Philosophy, some physicists and philosophers are still committed to providing a mechanistic explanation (or understanding) of QM.

9.3.2 *Still, the QM’s Challenges*

According to Kuhlmann (2018) there are other non-classical features, besides entanglement, where QM seems to clash with the ontological commitments of the New Mechanicism, namely the Indeterminacy of properties and Non-localizability of quantum objects. However, that is also a matter that falls on the ongoing discussion in the context of QM’s interpretations debate. For instance, according to Allori (2015), a particle position can be taken as a primitive variable, and therefore there is at least one property with always a value well-defined: position. Again, we could

say that what drives those interpretations is a reaction against QM's "weirdness" in the line of the ontological model based on mechanical philosophy. In that case, the situation in physics is not so different from what it was at the beginning of the eighteenth century.

On the other hand, Kuhlman argues that some of these features can be, in part, addressed by decoherence: at the "quantum level" that clash will still exist, but at the "classic level", due to decoherence, it could be almost unnoticeable. Nonetheless – of course – at the ontological level, this does not solve the incompatibility.

9.4 Against the Universality Thesis of QM

Even if we accept that QM is incompatible with (New) Mechanical Philosophy, why would it place a problem to New Mechanism in special sciences?

Kuhlmann provides one main reason: the universality of QM. That is, QM applies to all physical domains, and all physical phenomena should be grounded and explained ideally upon QM. Let us call it the universality thesis of QM. That is the reason why, even though decoherence can give an approximative explanation of why the "classic level" seems to be different from the "quantum level", it does not solve the problem of how to bridge the two realms since, in the end, everything is part of QM's domain. The assumption that the properties of the upper compositional level are supervenient, reducible or identical to the properties of the fundamental level (those that are the object of fundamental physical theories) are very present in von Neumann-Dirac axiomatisation and in most of the QM's debates. It is an atomistic heritage to assume that fundamental physical theories apply to all physical domains, or that, in principle, all physical reality is describable by fundamental physics.

However, why should we assume that all physical domains of reality are reducible to the most basic set of properties, relations and laws of the putative ultimate physical domain? Why don't we consider the possibility that classic objects have (some) different properties from quantum objects and endorse a pluralism ontology against the QM universality thesis? That is, to defend that there are ontologically emergent classical properties distinct and autonomous from quantum object's properties - for instance, the property of position. Therefore, we could accept that quantum objects do not have nor position or trajectory, but that would be unproblematic to any ontological account of all non-quantic domain.

It can be argued that QM does not contain any precise criterion for identifying the frontier between micro and macro or between quantum and classic. That there is nothing in QM fixing such a border. Nevertheless, on the one hand, since classic objects do not share the same set of properties as quantum objects, then is unsurprising that QM does not fix such a frontier. On the other hand, that border is not settled on a specific spatial scale since the claim is that (some) classical properties ontologically differ from quantum properties. That is, if classical properties ontologically emerge from quantum properties, the distinction between quantum and classic

(that is, ontological) is not identical to the difference between macro and micro (that is phenomenological). It, therefore, does not make sense to ask for a scale frontier. Also, it is essential to recall that at the macro-scale there are some quantum phenomena.

How to make sense of this emergentist hypothesis?

If we move towards a relational ontology according to which there are changeable structures of relations, then individuals are relational entities that can have qualitative transformations. The central assumption of this ontology is that every physical object is a relational entity occupying a relational location in a structural complex, and new levels of organisation or structure can emerge – against the one-level micro-physicalist picture of the world. Adopting this view, one is in a position to argue that a given structure can instantiate a new type of property that is not manifested at the level of the structure's components, i.e. the relations and their relata. Furthermore, one can take a step forward and try to explain the emergence of a structure's new property and its micro-irreducibility. This can be done by virtue of a qualitative change at the level of those objects as relata of the relations that compose the structure. That is, emergent structural attributes are attributes of specific macro-structured networks of transformative and interdependent relations between integrated system's parts. This can and must be explained via a relational-transformative account of interlevel emergence (Santos, 2015, 2021) as specific modes of the composition of their parts' transformative and structurally interdependent relations.

In this view, new emergent higher-level structural properties and laws may be generated from the transformative relations between lower-level sub-structures. Relations and structures could still be the actual star performers of science and reality, but a structuralist view of the world would not be equal to an essentially flat or static view of it. In particular, this view could be made possible if structures were seen as primarily 'concrete structures', taken as relations between first-order properties, in contrast to mere 'abstract structures', taken as higher-order, formal (logical or mathematical) properties of relations (Cordovil et al., 2022).

Within this ontological working hypothesis in a measurement there is a qualitative change of the quantum object/system being measured in virtue of the new relational network that integrates (quantum object/system – measurement object/system). Namely, since the measurement object/system is characterised by having the property of position, then every other object/system can only be a relatum of the measurement relation if it undergoes a qualitative change from which emerges the property of position. This qualitative change is not reducible to its microstructure. That is, we can think that, in essence, a measurement is an act of transformation of physical reality.

On the other hand, this ontological proposal would understand the role of decoherence as the interaction between the quantum object's structure and the environment.

Moreover, this view seems to go precisely in the direction of the mechanistic claim that levels are not monolithic stratification of the universe, nor are they fundamentally a matter of size or causal interactions within a level.

Besides, if it is vital to New Mechanical Philosophy to reject the idea of universal laws that grounds, axiomatically, all explanations in science, then the rejection of the universality thesis of QM is the more natural move to make.

9.5 Conclusion

In conclusion, as in the late seventeenth century, contemporary fundamental physics (or at least, QM) is in contradiction with mechanistic ontology. However, the reaction to the mystery placed by the gravitational force in Newton, like the reaction to some mysterious quantum features, seems to have in common the (spontaneous) defence of some kind of mechanical ontology. So, in fact, the situation does not seem to be new and if Physics has not been playing a central role in the literature devoted to New Mechanical Philosophy that may not be simply a consequence of the hypothetical incompatibility between QM and Mechanicist ontology, but due to adhesion to a specific QM's interpretation and the micro-physicalist assumption of the universal character of QM.

Acknowledgements I acknowledge the financial support of FCT, 'Fundação para a Ciência e a Tecnologia, I.P.' under the Stimulus of Scientific Employment - DL57/2016/CP1479/CT0065.

References

- Allori, V. (2015). Primitive ontology in a nutshell. *International Journal of Quantum Foundations*, 1(3), 107–122.
- Araújo, J. E. F., Cordovil, J. L., Croca, J. R., Moreira, R., & da Silva, A. R. (2009). A causal and local interpretation of experimental realization of Wheeler's delayed-choice Gedanken experiment. *Apeiron*, 16(2), 179–190.
- Bohm, D. (1951). *Quantum theory*. Prentice-Hall.
- Cohen, B. (1999). A guide to Newton's principia. In *The principia: Mathematical principles of natural philosophy: A new translation by I. Bernard Cohen and Anne Whitman* (pp. 11–370). University of California Press.
- Cordovil, J. L. (2015). Contemporary quantum physics metaphysical challenge: Looking for a relational metaphysics. *Axiomathes*, 25, 133–143. <https://doi.org/10.1007/s10516-014-9259-2>
- Cordovil, J. L., Santos, G. C., & Symons, J. (2022). Reconciling ontic structural realism and ontological emergence. *Foundations of Science*, 28, 1–20. <https://doi.org/10.1007/s10699-021-09828-8>
- Craver, C., & Tabery, J. (2019). Mechanisms in science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 ed.). <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>
- Esfeld, M., Deckert, D., & Oldofredi, A. (2015). What is matter? The fundamental ontology of atomism and structural realism. In A. Ijjas & B. Loewer (Eds.), *A guide to the philosophy of cosmology*. Oxford University Press.
- Garber. (2013). Remarks on the pre-history of the mechanical philosophy. In D. Garber & S. Roux (Eds.), *The mechanization of natural philosophy* (pp. 3–26). Springer.
- Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

- Huang, K. (2007). *Fundamental forces of nature*. World Scientific Publishing.
- Janiak, A. (2008). *Newton as a philosopher*. Cambridge University Press.
- Janiak, A. (2014). *Newton: Philosophical writings* (revised ed.). Cambridge University Press.
- Janiak, A. (2021). Newton's philosophy. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2021 ed.) <https://plato.stanford.edu/archives/fall2021/entries/newton-philosophy/>
- Kuhlmann, M. (2018). Mechanisms in physics. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 283–295).
- Kuhlmann, M., & Glennan, S. (2014). On the relation between quantum mechanical and neo-mechanistic ontologies and explanatory strategies. *European Journal for Philosophy of Science*, 4, 337–359.
- Lopez, C. (2016). A local interpretation of quantum mechanics. *Foundations of Physics*, 46, 484–504. <https://doi.org/10.1007/s10701-015-9976-4>
- McGuire, J. E. (1972). Boyle's conception of nature. *Journal of the History of Ideas*, 33(4), 523–542.
- Newton, I. (1999). *The principia: Mathematical principles of natural philosophy: The authoritative translation by I. Bernard Cohen and Anne Whitman*. University of California Press.
- Santos, G. (2015). Ontological emergence: How is that possible? Towards a new relational ontology. *Foundations of Science*, 20(4), 429–446.
- Santos, G. (2021). Integrated-structure emergence and its mechanistic explanation. *Synthese*, 198, 8687–8711. <https://doi.org/10.1007/s11229-020-02594-3>
- Schrödinger, E. (1935). Discussion of probability relations between separate systems. *Proceedings of the Cambridge Philosophical Society*, 31, 555–563.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 10

Mechanistic Explanations in Physics: History, Scope, and Limits



Brigitte Falkenburg

Abstract Despite the scientific revolutions of the twentieth century, mechanistic explanations show a striking methodological continuity from early modern science to current scientific practice. They are rooted in the traditional method of analysis and synthesis, which was the background of Galileo's resolutive-compositive method and Newton's method of deduction from the phenomena. In early modern science as well as in current scientific practice, analysis aims at tracking back from the phenomena to the principles, i.e., from wholes to parts, and from effects to causes. Vice versa, synthesis aims at explaining the phenomena from the parts and their interactions. Today, mechanistic explanations are atomistic in a generalized sense. They have in common to explain higher-level phenomena in terms of lower-level components and their causal actions or activities. In quantum physics, the lower-level components are subatomic particles, and the causes are their quantum interactions. After the quantum revolution, the approach continues to work in terms of the sum rules which hold for conserved properties of the parts and the whole. My paper focuses on the successes and limitations of this approach, with a side glance at the recent generalization of mechanistic explanations in cognitive neuroscience.

Keywords Mechanistic explanation · Method of analysis and synthesis · Resolutive-compositive method · Aristotelian tradition · Atomism · Conserved properties · Cognitive neuroscience

10.1 Introduction

The successes of natural science are based on the experimental method and the mathematical models of Galileo's and Newton's physics. Closely related to the foundation of modern physics was the mechanistic world view, according to which all material bodies were conceived to consist of mechanical corpuscles or atoms and

B. Falkenburg (✉)
Technischen Universität Dortmund, Dortmund, Germany
e-mail: brigitte.falkenburg@tu-dortmund.de

to obey the laws of classical mechanics. Mechanistic thinking dominated the understanding of nature until the end of the nineteenth century, but with the advent of modern atomic physics, it became apparent that a mechanistic understanding of nature in terms of classical physics was incompatible with the atomistic structure of matter and the interactions of subatomic particles. Hence, the mechanistic approach to nature was generalised in twentieth century physics, chemistry, biology, and the investigation of neuronal mechanisms in neuroscience.

However, we should be careful about what “mechanistic” still means today. In the philosophy of biology and neurobiology, the roots of mechanistic explanations in early modern science and their scope in current scientific practice are usually not discussed. In view of the successes of cognitive neuroscience, it has been claimed that a complete scientific explanation of the world, including human consciousness, is in principle possible. Several neuroscientists and philosophers supported a deterministic world view according to which the human mind reduces to the neural mechanisms in the neocortex and free will is an illusion generated by the brain (Churchland, 1995; Roth, 2003; Singer, 2003, 2004; Rubia, 2009). In Germany, this view gave rise to a heated public debate, to the point of calling for changes in criminal law (Geyer, 2004). In the last decade, it became clearer that the human brain is tremendously complex and that the mechanistic explanations of neuroscience are not as far-reaching as expected. Hence it is time to take a step back and clarify their meaning and scope.

Despite the scientific revolutions of the twentieth century, mechanistic explanations show a striking methodological continuity from early modern science to current scientific practice. They are rooted in the traditional method of analysis and synthesis, which was the background of Galileo’s resolutive-compositive method and Newton’s method of deduction from the phenomena. In early modern science as well as in current scientific practice, analysis aims at tracking back from the phenomena to the principles, i.e., from wholes to parts, and from effects to causes. Vice versa, synthesis aims at explaining the phenomena from the parts and their interactions. Today, many mechanistic explanations are atomistic in a generalized sense. They have in common to explain higher-level phenomena in terms of lower-level components and their causal actions or activities. In quantum physics, the lower-level components are subatomic particles, and the causes are their quantum interactions. After the quantum revolution, the approach continues to work in terms of the sum rules which hold for conserved properties of the parts and the whole. My paper focuses on the successes and limitations of this approach, with a side glance at the recent generalization of mechanistic explanations in cognitive neuroscience.

10.2 The Origin of Mechanistic Explanations

The philosophical background of mechanistic explanations is the mechanistic world view of early modern science and philosophy, according to which all natural processes were considered as mechanisms, or to function like machines. The proponents of

this world view were Descartes and Hobbes, regardless of their profound philosophical differences concerning dualism or materialism. However, the mechanistic world view is much older. The founders of early modern science and philosophy took up ancient atomism. Another crucial mechanistic paradigm was to compare the solar system with a clock, as illustrated by the famous astronomical clocks in European cathedrals since the Late Middle Age. Indeed, important mechanical explanations in early modern science relied on the analogy between the universe and a clock. Later, the laws of Newton’s mechanics explained the dynamic structure of the celestial clock, or the machinery of the universe, in terms of gravitation as a universal force.

In general, a mechanism is a causal structure, or more precisely, a system of elements that work together to bring about or cause a process. The English term ‘mechanism’ derives from the Greek word *μηχανή* for ‘machine’ and the corresponding Latin word *mechanica* or its derivatives. A British dictionary defines a mechanism primarily as “1. a system or structure of moving parts that performs some function, especially in a machine” (Collins, 2012). A simple example of a mechanism is a clock. The system of elements is the clock. Its causal elements are the balance and the gears, which interlock in such a way that they move the clock hands to indicate the time on the dial (Fig. 10.1).

A mechanistic explanation, then, explains a phenomenon or process by a mechanism, i.e., in terms of certain causal elements or components that work together like the parts of a machine. The above dictionary extends the explication of a mechanism to this analogous use and gives a second definition, according to which a mechanism is “2. something resembling a machine in the arrangement and working of its parts: the mechanism of the ear” (Collins, 2012). This second, analogical meaning of the term ‘mechanism’ and its extension to the way in which organs function also emerged in the seventeenth century and dates to ancient atomism. Even Aristotle discussed the analogy between technical tools or machines and the processes of nature (Aristotle, *Physics*, 199a), albeit within his anti-atomistic, teleological account of nature. Early modern science dispensed with Aristotle’s teleological explanations of mechanical processes to explain vice versa the way in which organs function in mechanical terms. In the Renaissance,

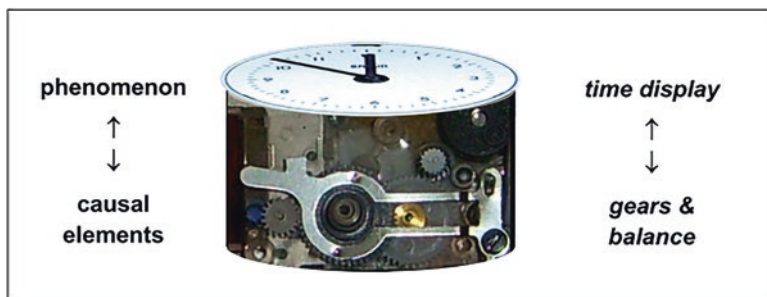


Fig. 10.1 The clock, a simple mechanism

mechanical analogies entered medical science. The title of Andreas Vesalius' famous anatomy textbook *De humani corporis fabrica* (Vesalius, 1543) paradigmatically expresses the analogy between the structure of the human body and an artificial structure. After the reception of the Arabian theory of vision in the Late Middle Age, and the rediscovery of ancient atomism as described by Lucretius (1570), early modern science started to develop mechanistic explanations of sensory perception (Hobbes, 1655).

10.2.1 *The Tradition of Analysis and Synthesis*

The mechanistic explanations of early modern science and philosophy were grounded in the ancient method of analysis and synthesis, an inductive method which has remained influential up to current scientific practice. It gives rise to a generalized mechanistic methodology, which is typical of the “dissecting” sciences (Schurz, 2014, 35) from Galileo's and Newton's days to recent neuroscience. The Greek terms *analysis* and *synthesis* mean “decomposition” and “composition” respectively. Today, exactly this meaning is still found in the practice of chemical analysis and synthesis. However, the traditional analytic-synthetic method of the exact sciences is much more complex. In the accounts of Galileo or Newton, analysis combines *decomposition* and *causal analysis*, whereas synthesis vice versa combines (re-) *composition* and *causal explanation*.

If we look at the typical structure of a mechanism (Fig. 10.2), we see that the analysis proceeds *top-down* from the whole to its parts, from the phenomenon or process to be explained down to the entities and interactions that compose it. The synthesis runs vice versa; it proceeds *bottom-up* from the parts and their interactions to the whole and from these entities and interactions as the *explanans* to the phenomenon or process *explanandum*. In modern philosophy, the analysis is inductive (or abductive). It gives rise to an inference to the causal structure that underlies a

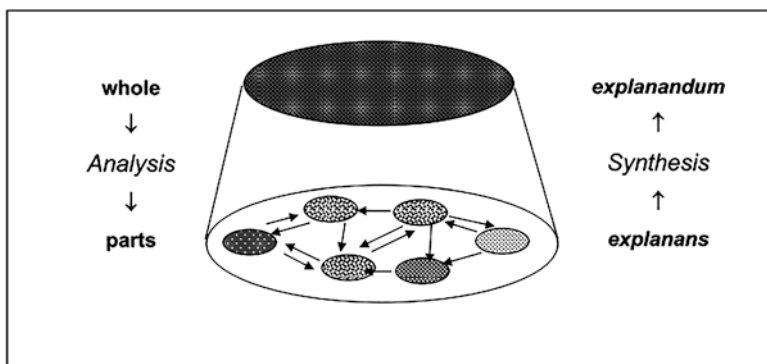


Fig. 10.2 Typical structure of a mechanism

phenomenon, i.e., it aims at an interference to the best explanation. The synthesis is then the corresponding mechanistic explanation of the phenomenon in terms of lower-level causal entities.

The method of analysis and synthesis traces back to ancient geometry and medicine, and it was widely shared in early modern science and philosophy (for details, cf. Beaney, 2021). However, there were two different methodological traditions of analysis and synthesis which merged in early modern science. On the one hand, the Aristotelian tradition of Latin medieval science and philosophy developed a resolutive-compositive method, *resolutio* and *compositio* being the Latin terms for *analysis* and *synthesis*. The medieval resolutive-compositive method combined inductive and deductive elements of reasoning. It remained attached to Aristotle's conception of induction and deduction and proceeded in terms of the logical connections between antecedents and consequences. In empirical science, these logical relations became associated with relations of cause and effect. Hence, the resolutive part of the method was the regress to causes, whereas the compositive part was a causal explanation. In the thirteenth century, this method was advanced by Robert Grosseteste, and later, in the Padua school of early modern science, by Giacomo Zabarella. Their method resembled Galileo's methodology (Crombie, 1953), but they still adhered the Aristotelian tradition of rejecting any mathematical analysis of the phenomena (Engfer, 1982, 95; Hintikka & Remes, 1974, 107–108).

On the other hand, there was the analytic-synthetic method of ancient geometry explained in Pappus's commentary on Euclid's works. In medieval science and philosophy, it was not available for a long time. Via the Arabic tradition, Pappus's method partially received in geometrical optics. Alhazen's Book of Optics (*De Aspectibus or Perspectiva*) was translated into Latin around 1200. Witelo's influential *Perspectiva* written around 1270 built on it (Lindberg, 1971, 1976; Crombie, 1953) and it refers to some proofs and geometrical constructions that seem to stem from a Latin partial translation of Pappus's commentary (Unguru, 1974). Pappus's complete Greek text and its Latin translation became only generally accessible in the Renaissance (Pappus, 1589; translation: Hintikka & Remes, 1974, 8–9). Then, it became very influential in the mathematical tradition of early modern science and philosophy.

Pappus's method of geometry was an inductive procedure that substantially differs from induction in the Aristotelian sense. For Pappus, analysis and synthesis were the complementary parts or steps of a joint regressive-progressive method. Its "analytic" part is regressive, it infers from something that is taken for given to an underlying first principle by running through the antecedents of this given (or assumed) consequence (Hintikka & Remes, 1974, 11–14). Its second, "synthetic" part is progressive or deductive. It aims at confirming the principles established by analysis by deriving from them what was originally given or assumed, i.e., by proceeding from the principle found by analysis to its consequences. So far it resembled the resolutive-compositive method of the medieval Aristotelian tradition, however, with two crucial differences: *first*, the inferences from consequences to antecedents or vice versa employed geometrical constructions; *second*, analysis was only the first part of the method, which was to be completed by synthesis.

In contrast to their medieval predecessors, Galileo and Newton used geometrical constructions to analyse the phenomena into their causal components, resorting to Pappus's account of analysis and synthesis and distinguishing their new science in this way substantially against the earlier scientific traditions. In addition, however, they adopted the causal aspect of the medieval resolutive-compositive method of the Aristotelian tradition. Their mathematical and experimental analysis of the phenomena into idealized components came along with causal analysis. Both reinterpreted the logical relations of antecedents and consequences of the resolutive-compositive method in terms of physical causes and effects; and they combined these causal relations with a mathematical analysis of the phenomena into components, which then are investigated by the experimental method. The resulting combined method of analysis and synthesis establishes a complex pattern of part-whole relations and causal relations. The causal structure of this complex pattern is investigated by mathematical and experimental analysis. This combined analytic-synthetic method and its application in experiments serves to analyse, explain, and predict natural phenomena in mathematical terms. Such a complex pattern of part-whole relations and causal relations is indeed typical of natural science up to the present day, and it corresponds to the structure of the mechanistic explanations on which recent philosophy of science focuses.

10.2.2 *Newton's Methodology*

Galileo did not explain his resolutive-compositive method in philosophical terms, but he practised it in his famous experiments with the inclined plane, in which he changed the inclination to analyse the causal components of the falling motion (Losee, 1993). Newton's main works, however, contain several methodological considerations. Roger Cotes relied on them in the preface to the second edition of Newton's *Principia*, where he stated that natural science proceeds.

according to a twofold method, the analytical and the synthetic. They derive the forces of nature and their simple laws from a few selected phenomena by means of analysis, and present the former, by means of synthesis, as the nature of the remaining phenomena. (Cotes, 1713, 386).

This remark is very similar to Newton's account of the analytic-synthetic method in *Query 31* of the *Opticks*. There, he compares the method of natural science to the corresponding method of mathematics. Following Pappus, he emphasizes that the analysis or decomposition has always to be performed before the synthesis or composition:

As in Mathematicks, so in Natural Philosophy, the Investigation of Difficult Things by the Method of Analysis, ought ever to precede the Method of Composition. (Newton, 1730, 404)

Then, he emphasizes that in physics the method of analysis and synthesis establishes part-whole relations and causal relations. In this way, he brings Pappus's geometrical method together with the Aristotelian tradition of the resolutive-compositive method and reinterprets the latter in terms of physical causes and effects. The analysis proceeds from phenomena to their components and causes; the effects in nature are motions; their analysis aims at finding the forces that cause them. The synthesis, conversely, serves to prove that these causes can indeed explain the phenomena:

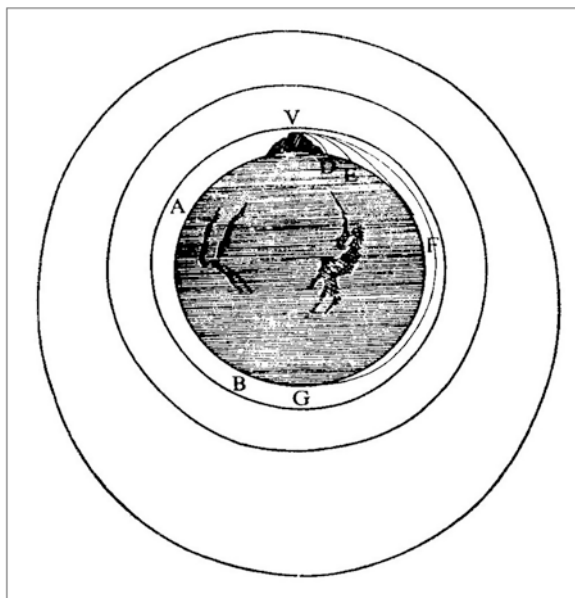
By this way of Analysis we may proceed from Compounds to Ingredients, from Effects to their Causes, and from Motions to the Forces producing them; and in general, from Effects to their Causes, and from particular Causes to more general ones, till the Argument end in the most general. This is the Method of Analysis: And the Method of Synthesis in assuming the Causes discover'd, and establish'd as Principles, and by them explaining the Phaenomena proceeding from them, and proving the Explanations. (Newton, 1730, 404).

These remarks on the analytic-synthetic method in the *Opticks* and the rules of philosophizing in the *Principia* point to the same method of “deduction from the phenomena” (Achinstein, 1991, 32–50; Worrall, 2000). At the beginning of *Book III* of the *Principia*, Newton gives four methodological rules to explain the analytic method (Newton, 1726, 794–796). The first two refer to the causal analysis of phenomena. They demand that no more causes be assumed than are sufficient to explain the phenomena, and that similar effects be attributed to similar causes. The third is a rule of induction, which demands to generalise the empirically known mechanical properties of bodies to *all bodies*, including the smallest constituents of bodies, i.e., the atoms, which Newton thought to exist. The fourth rule demands that empirically established hypotheses be maintained unless they are falsified, instead of considering contrary speculative hypotheses. This rule expresses a pragmatic conservatism concerning well-established theories. It not only conforms to Newton's famous dictum *hypotheses non fingo* (Newton ...), but also to the following remark in his *Opticks*:

This analysis consists in drawing general conclusions from experiments and observations by induction, and in admitting no objections to them which are not taken from experiments or from other certain truths. For hypotheses are not considered in the experimental study of nature. (Newton, 1730, 404).

Hence, analysis in Newton's sense combines the dissection of phenomena into components or of bodies into their constituent parts (third rule) with causal analysis (first and second rule). For the conclusions drawn from the phenomena, experimental observations are the touchstone (fourth rule and the passage just quoted). In the *Principia*, Newton shows that the analysis of the phenomena according to his rules of philosophizing gives rise to an explanation of the trajectory of thrown mechanical bodies on earth and the motions of celestial bodies in terms of one and the same cause, gravitation. A diagram in the appendix to the *Principia* demonstrates that there is a continuous transition from Galileo's parabola of the motion of a thrown body to the Kepler orbit of the moon around the earth, in accordance with Newton's first and second rules (Newton, 1729, 551; Fig. 10.3).

Fig. 10.3 Transition from Galileo's to Keplerian motion (Newton, 1729, 551)



The rules of philosophizing correspond to the analytic step of the analytic-synthetic method, while the axiomatic approach of the *Principia* in terms of definitions, laws of motions, and mathematical deductions corresponds to the synthetic step. The synthesis is the mathematical deduction of the motions from the law of force and gravitation. Only the latter step corresponds to the deductive-nomological (DN-) account of scientific explanation which dominated the philosophy of science for such a long time. For the *Opticks*, no such synthetic step from the principles of an atomistic theory of light to a deduction of the optical phenomena in terms of a mechanistic explanation was available to Newton. Here, he demonstrated the interplay of analysis and synthesis only by the experimental decomposition of white light into the spectral colours and by the opposite composition of white light from the coloured light rays by the superposition of two spectra of prisms arranged in parallel, which in turn yield white light (Fig. 10.4).

In Newton's days the axiomatic, or synthetic, approach corresponding to DN-explanations only worked for mechanics as a full-fledged mathematical theory of gravitation and the mechanical motions of bodies. But it did not work for Newton's optics. In this field, Newton was not able to support his analytic inference to light atoms as the best explanation of the phenomena discussed in the *Queries* of the *Opticks* by an atomistic theory of matter and light. Such a theory was not only beyond the scope of Newton's optics, but also of nineteenth century physics. With the rise of quantum theory, it turned out that mechanistic explanations based on the laws of classical physics cannot cope with the atomistic structure of light and matter.

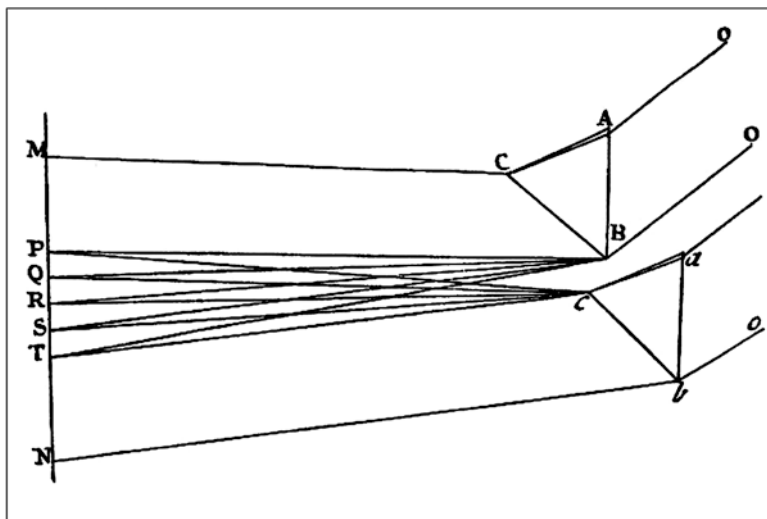


Fig. 10.4 Analysis and synthesis of light (Newton, 1730, 147)

10.3 Mechanistic Explanations Today

In twentieth century physics and beyond, mechanistic explanations were generalized in an inflationary way. To talk of mechanisms is ubiquitous in science and technology today. One speaks of the mechanism of the steam engine, the mechanism of signal transmission through light or radio waves, the electrodynamic and thermic mechanisms of the formation of a thunderstorm, the astrophysical mechanisms of generating cosmic rays, etc., and even the Higgs mechanism of the standard model of particle physics that explains the mass of subatomic particles. Examples from biology are the mechanisms of photosynthesis, of the replication of DNA, or of gene expression. We may add examples from neuroscience, above all the neural mechanisms that explain pattern recognition and learning by neural networks. In all these cases (except the Higgs mechanism, I suspect), these mechanisms explain a certain phenomenon or process in terms of part-whole relations and causal relations. Combining part-whole relations and causal relations, they retain the crucial features of the mechanical explanations of early modern science, i.e., they reproduce the inferential and explanatory structure of Galileo's or Newton's analytic-synthetic method. In addition, the mechanisms on which they rely still draw on the old analogy between processes in nature and the mechanisms of machines. The mechanisms of the steam engine, the formation of a thunderstorm, the generation of cosmic rays, photosynthesis, DNA reduplication, etc., including neural mechanisms, all have in common that they produce a phenomenon or process by their moving parts, or, generally, by the dynamics of their elements.

10.3.1 *The Recent Philosophical Definitions*

At this point we may look at the definitions of a mechanism in recent philosophy of science. The proponents of the recent “mechanistic turn” in the philosophy of science emphasize this dynamic aspect, but in quite different regards. Wesley Salmon and Stuart S. Glennan define the concept of mechanism in terms of causal processes or causal laws, having the laws of twentieth century physics in mind. Salmon (1984, 240) emphasizes that an adequate account of scientific explanation requires mechanistic explanations in a generalized sense and that they may even employ fields (*ibid.*, 241). According to him, a mechanism is any causal fork or causal process, including stochastic processes:

The theory here proposed appeals to causal forks and causal processes; these are, if I am right, the mechanisms of causal production and causal propagation that operate in our universe. These mechanisms [...] may operate in ineluctably stochastic ways. (*ibid.*, 239).

Hence, Salmon identifies mechanistic explanations and causal explanations. In his 1984 book *Scientific Explanation and the Causal Structure of the World*, he defined causal processes in terms of mark transmission, but later, in terms of the transmission of a conserved quantity between two events (Salmon, 1997). According to both definitions, the paradigm case of a causal process is signal transmission in physics, such as the emission, propagation, and detection of radio waves or light signals, including quantum processes such as the emission, propagation, and absorption of photons. This conception of a mechanism is very general. It holds for the classical impact of two billiard balls as well as for the transmission of a quantum signal, which obeys the principle of energy conservation and the probabilistic laws of a quantum theory. In addition, it is very basic. Signal transmission is a causal process that propagates from cause A to effect B, where A and B belong to one-and-the-same level of phenomena. The part-whole relations, which are crucial for Newton’s analytic-synthetic method and the mechanistic explanations of early modern science, are missing here. According to Glennan, this approach indeed is *too* basic. According to Glennan’s definition, a mechanism is a complex system with causal components that interact via causal laws, i.e., it involves part-whole relations:

A mechanism underlying a behavior is a complex system which produces that behavior by [...] the interaction of a number of parts according to direct causal laws. (Glennan, 1996, 52).

Salmon’s and Glennan’s definitions are restricted to mechanisms in physics. Both definitions fall short of a *general* concept of causality. Causation *in general* cannot be reduced to mechanistic causations, as the case of causation by omission shows (according to Dowe, 2008). Glennan takes the opposite route: he attempts to reduce mechanistic explanations to causal laws. His goal is to explain higher-level causal processes in terms of lower-level laws, whereas the fundamental laws of physics that explain the higher-level processes according to him are not subject to mechanistic explanation. However, Glennan’s attempt to reduce mechanistic explanations to causal laws does also not work, insofar as it neglects the crucial part-whole relations supported by the causal laws.

In physics, the causal processes underlying a mechanism are often described in terms of a physical dynamics. The solar system, as the paradigm case of a mechanism of physics, is a complex bound system of bodies, with gravitation as the binding force that keeps the planets and moons in their orbits around the sun or the planets. In quantum mechanics, this example of a classical bound system of mechanical bodies has been generalized as follows. The atoms are described as bound systems of charged particles, i.e., an atomic nucleus plus N electrons described by an N -particle quantum mechanical wave function. The electrons within an atom have no orbits, but they are kept in bound quantum states via the Coulomb force; the underlying causal law is the N -particle Schrödinger equation. Analogously, the atomic nucleus is an N -particle system of protons and neutrons which are kept together according to the quantum laws of the strong interaction.

In biology, this approach is possible to the extent that chemical or biochemical mechanisms are at work, which in turn reduce to the mechanisms of molecular physics, in biophysics and genetics (cf. Odenbaugh & Griffiths, 2022). An example of applying the laws of physics to biophysical processes in a mechanistic explanation is the computer simulation of protein folding, which however is still a very complex problem without solution (Vallejos & Vecchi, Chap. 6, this volume). In cell biology, epigenetics, evolution theory, neurobiology, etc., the situation is also very complex and difficult. In many cases, the causal laws available to explain the way in which the causal components of a mechanism work are laws in a very weak sense. Or no causal laws at all are known, as in the case of the heuristic assumption of mental mechanisms (Bechtel, 2008). Therefore, philosophers of biology and neuroscience typically define a mechanism without recourse to causal laws:

Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. [...] Activities are the causal components in mechanisms. Mechanisms are composed of both entities (with their properties) and activities. Activities are the producers of change. Entities are the things that engage in activities. Activities usually require that entities have specific types of properties. (Machamer et al., 2000, 3.)

Activities are the causal components in mechanisms. [...] mechanisms are entities and activities organized such that they exhibit the explanandum phenomenon. (Craver, 2007, 6).

A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism is responsible for one or more phenomena. (Bechtel & Abrahamsen, 2005, 423).

These definitions strikingly resemble the above dictionary definitions quoted above (Collins, 2012). According to them, mechanisms explain higher-level phenomena in terms of lower-level causal components, in a “dualistic” account of the components and their causal activities (Schiemann, 2019). Such explanations do not specify which kinds of causal activities are at work and how they relate to the causal entities. They just rely on the analogy between a complex system in nature and a machine with well-defined moving parts.

10.3.2 *Causal Components and Their Dynamic Properties*

Much philosophical confusion about the legitimacy of generalized mechanistic explanations arose from taking the mechanical analogy with the moving parts of machines too literally. To clarify the limits of this analogy a look at physics is helpful. For the case of physics, it is easy to specify the causal entities and their activities in precise terms, i.e., in terms of a physical dynamics. On this basis it is also easy to see how the concept of a mechanism can be generalized accounting for the physics after Newton, from electrodynamics to quantum mechanics and quantum field theory.

In contrast to the machinery of the gears inside a clock, the spatial structure of the parts of a mechanism is not necessarily decisive for the way it works, as the case of physics shows. Already William S. Malisoff (1940) made this point in defense of generalized mechanistic explanations. Indeed, the views about the mechanisms of nature in classical physics rely on reinterpreting the moving parts of mechanical machines in terms of idealized mathematical entities, such as the point masses and forces of mathematical physics:

What did the physicists of 70 years ago speculate about? I should say they speculated about mechanism itself. What is a mechanism? [...] A mechanism, they thought, is essentially a machine. And what is a machine? Simply enough, [...] a thing of cogs and levers. (Malisoff, 1940, 405)

The difference, however, between a physicist and a machinist was that the physicist's cogs and levers and machines consisted of mathematical points, lines surfaces, volumes, interacting by a system of forces between the points to which were attributed masses and velocities. (ibid., 405–406)

This observation perfectly agrees with Newton's account of the analytic-synthetic method of early modern science. Above all, it holds for the mechanism of the solar system. Classical mechanics replaces the celestial bodies by point masses, given that their extension is negligible as compared to the distance between them. It describes the causal properties of physical systems in terms of dynamic magnitudes such as mass or charge. For Malisoff, it is therefore obvious how to generalize the traditional mechanical physics to an up-to-date version of mechanistic explanations, in the age of relativity and quantum theory:

Do we not still use forces, particles, and the like, where we can? (ibid., 414).

Leaving aside the role of idealizations in physics, another argument results in the same conclusion. A mechanistic explanation explains the functioning of a mechanism, or a machine, in terms of its causal components. A purely spatial interpretation of the part-whole relationship of a mechanism does not match the way the moving parts of a machine function. The relevant part-whole relation primarily concerns the causal properties, not the spatio-temporal properties of the components of a machine or a mechanism. It is a relation between the causal or dynamic properties of the whole on the one hand and its parts on the other. The parts of a mechanism act as dynamic parts. Their causal activities correspond to their dynamic properties.

Philosophers may object that the expressions “causal parts” and “dynamic properties” are unclear and much debated. Current philosophy spells dynamic properties out in terms of dispositions, and the concepts of causality range from various successors of Hume’s regularity theory over variants of Salmon’s physics-based approach to Woodward’s interventionist account. But this should not worry us here. To cope with the mechanistic explanations in the practice of physics, the philosophical debates on dispositions and causality may be left aside. Instead, the above-mentioned physics-based approaches to causal laws and processes (Salmon, 1984, 1997; Glennan, 1996) matter, and, in addition, the part-whole relations that are constitutive for the dynamics of the compound systems described by physics.

A difficulty of relying on the causal laws and processes of physics is that different physical laws and theories give rise to several accounts of causality, from Einstein causality, i.e., the deterministic transmission of a physical signal within the light cone, to the irreversible processes that cause an entropy increase, or the indeterministic effects of a quantum measurement. Up to now, there is no unambiguous, well-established concept of causality or theory of causal processes that the physics community would share. There is no unified theory of physics, and the diverse concepts of causality used in the context of different theories cannot be unified either. But this should also not worry us here. To understand the mechanistic explanations of physics, we do not need a unified theory of physics but only models of specific physical phenomena with a well-defined underlying dynamic.

Beyond classical mechanics, *any* physical dynamics may give rise to mechanistic explanations in a generalized sense, from electrodynamics to thermodynamics, quantum mechanics, quantum field theory, or general relativity. Salmon (1984, 241) emphasized that mechanistic explanations in a generalized sense may even employ fields. The above case of the quantum mechanical description of atoms also shows that in general the parts of a mechanism do not need to be local, spatially well-identified parts. Any physical dynamics expresses the causal part-whole structure of a mechanism in terms of the dynamic properties of the whole and its constituent parts.

10.3.3 *The “Atomistic” Constitution of Matter*

From Newton’s mechanics to quantum physics, the dynamic properties of physical systems and their components are conserved physical quantities such as mass, charge, energy, and so on. Particles in a generalized sense are collections of such dynamic properties for which conservation laws hold. The quantum revolution dispensed with particle trajectories. What remained, however, is the concept of particles as collections of conserved quantities such as mass, charge, etc., which cause the hits and tracks in particle detectors (Falkenburg, 2007). This generalized particle concept corresponds to Eugene P. Wigner’s definition, according to which particles (or fields) are the irreducible representations of symmetry groups (Wigner, 1939). According to this most general particle concept, the relation between particles

and forces, or interactions, rests on the dynamic symmetries associated with conservation laws for mass-energy, charge, spin, parity, and so on.

In particle physics, these conservation laws and their experimental tests in high energy scattering experiments have been decisive for the quark-parton constituent model of protons and neutrons. They paved the way towards the standard model of particle physics. The dynamics of a compound quantum system gives rise to sum rules for the conserved quantities of subatomic particles and the complex quantum systems made up of them. The quantum parts of matter are defined in terms of sum rules for mass-energy, momentum, charge, spin, parity, etc., which are empirically tested in the scattering experiments of atomic, nuclear, and particle physics (Falkenburg, 2007, chapters 4 and 6). In nuclear physics, the binding energy of the protons and neutrons adds to the sum of the masses of protons and neutrons. In the quark model of particle physics, the situation is similar, but more complex, given that here also gluons and quark-antiquark pairs contribute to the energy of the matter constituents measured in the scattering experiments of high energy physics. Similar sum rules, however, hold for the number and kinds of the quasi-particles in a solid, which are investigated in condensed matter physics (Falkenburg, 2015); or for the strength of an electromagnetic field and the occupation number of the corresponding quantum field, that is, for the intensity of light and the expectation value of the number of photons in this quantum field. In all these examples, the quantum parts of matter and light are subject to a dynamic part-whole relation, instead of being spatial parts of matter or fields.

So, the causal components of mechanistic explanations in current physics are dynamic parts of matter or fields, that is, the dynamic parts of the N-particle quantum systems that constitute matter or the N-particle quantum states that make up fields. These dynamic parts of matter or fields are particles in a generalized sense. The corresponding part-whole relations are sum rules for conserved quantities such as mass-energy, charge, spin, parity, and so on. The resulting mechanistic explanations are atomistic in a general sense, i.e., they rely on the generalized particle concept of the current quantum theories and particle physics. The “atoms” of current physics are the subatomic particles that exist according to the standard model of particle physics. This observation supports the following definition of a mechanism in physics:

A mechanism is a complex system which produces a certain physical phenomenon by the interaction of a number of causal components with conserved dynamic properties that interact according to the laws of a physical dynamics and constitute the system as a whole in accordance with sum rules for the conserved quantities of this dynamics.

Here, the definitions suggested by Glennan (1996) and his followers are specified in terms of a physical dynamics, and Salmon’s (1997) account is generalized in such a way that it includes the compound systems of physics. This approach substantially differs from that of dynamical system theory (cf. Kaplan, 2018) by including not only the differential equations of a physical dynamics, but also the related conserved dynamic quantities, which in turn are the basis for establishing a dynamic part-whole relation between a complex system and its causal components.

It should be added that mechanisms in this sense do not only explain processes, i.e., phenomena of change. They can also explain under which conditions there is *no* change. The mechanisms of classical mechanics or quantum mechanics explain the *stability* of compound systems of bodies or subatomic particles. Newton's theory of gravitation explains the stability of the solar system in terms of the approximate Kepler orbits of the planets and moons. Quantum mechanics explains why and under which dynamic conditions atoms and atomic nuclei are stable.

Nevertheless, the mechanistic explanations of physics in this sense have crucial limits. They cannot cope with mechanisms based on classical continuum mechanics or thermodynamics (for examples, see Falkenburg, 2019, 85–87). A quantum field with well-defined phase, but unsharp occupation number is obviously beyond the scope of the above definition. To what extent mechanisms in this sense can explain collective behavior such as phase transitions is also unclear. Philip W. Anderson is famous for his essay *More is Different* which emphasises that complex systems have many non-reducible properties (Anderson, 1972). In his introductory textbook on solid-state physics, which explains, e.g., how quantum physics explains the magnetic properties of solids, he emphasizes at the beginning: “*We do not know why there are solids*” (Anderson, 1997, 3).

Even if there is no *complete* (quantum) explanation of why (classical) solids exist, however, many properties of solids can be explained by the dynamics of their subatomic constituents. Therefore, ontological reduction works in physics *top-down* from macroscopic bodies to molecules, atoms, electrons, atomic nuclei, protons, neutrons, and finally, quarks and gluons; and mechanistic explanations in the above generalized sense suggested here work *bottom-up* for the constitution of matter in terms of the dynamic properties of subatomic particles.

10.4 Mechanistic Explanations in Neuroscience, and Their Limits

So, what are generalized mechanistic explanations good for? The above-mentioned restrictions suggest that the definition suggested in Sect. 10.3.3 only works if the *number* of causal components of a mechanism is well-defined. Otherwise, to talk of a mechanism seems to be a mere *façon de parler*, since any analogy with the functioning of a machine fails. Two other obvious necessary conditions for a successful mechanistic explanation in the sense of Sect. 10.3.3 are that the *dynamic properties* of the causal components are known, and that it is possible to specify a *dynamic part-whole relation* that connects the properties of the complex system with the properties of its causal components. In physics, this part-whole relation is defined in terms of the sum rules that hold for mass-energy, momentum, charge, spin, parity, etc. To a large degree, this approach can also be generalized to the higher-level explanations of chemistry, biochemistry, molecular biology, and neurobiology. Many of these mechanisms work via electro-chemical signal transmission, for

which the conservation laws of charge and energy hold. Hence, their basis is a physical dynamics, as in the electric circuit model of signal conduction along the membrane of an axon (Hodgkin & Huxley, 1952) which is based on the laws of electrodynamic.

These considerations also shed light on the scope of mechanistic explanations in neuroscience. The theory of neural mechanisms is based on the Hodgkin-Huxley model just mentioned, the laws of chemical signal transmission through the synapses, and the theory of artificial neural networks. The theory of artificial neural networks describes the functioning of the parallel computers which underly the technological achievements of machine learning etc., which have gained increasing importance in all scientific disciplines and branches of technology during the last decades. Here, the analogy between processes in nature and the way a machine works runs in both directions: Artificial neural networks are modelled after the structure and functioning of neural networks; and, vice versa, the way the neural network in the brain functions is interpreted in terms of the functioning of a parallel computer. So far, so good. A parallel computer is a complex system, the causal components of its hardware function according to the laws of physics, and in this respect, it is a mechanism in the sense of Sect. 10.3.3. The phenomenon which this mechanism produces is the computer output of a calculation, and/or the way in which a robot moves according to the results of the calculation.

To compare the neural network in the brain with a computer is an important heuristic tool of computer science as well as neuroscience. The computer model of the brain is a highly idealized, strongly simplified, very crude model of the brain, given that the brain is the most complex system known in the universe. But planets, too, are no mass points; nor belong the atoms and their constituent parts to the laws of classical mechanics; and the computer model of the brain is no more wrong or less true than Newton's atomic model was. Even though the laws of classical mechanics failed in atomic physics, the classical atomic model of Rutherford, and its deficiencies, paved the way first to Bohr's atomic model, and then, to quantum mechanics.

However, the analogy between the brain and a computer crucially differs from Rutherford's analogy between the atom and the solar system, or from Newton's way of attributing the dynamic properties of mechanical bodies to the atoms, following his third, inductive, rule of philosophizing (Sect. 10.2.2). The celestial bodies in the solar system, the atoms, and the subatomic constituent parts of matter share the dynamic property of mass. The law of gravitation, the Coulomb law of electrodynamic, and the Schrödinger equation of the hydrogen atom predict compound N-body or N-particle systems which are bound together by the conserved dynamic properties of (gravitational) mass and electric charge. But no mechanism is known that might explain how the brain produces the conscious human mind. No dynamics is available for the relation between brain and mind, nor do the brain and the mind share any properties on which the analogy between the brain and a parallel computer may rely. To bring then "information" into play is more confusing than illuminating. The mathematical information processed by a computer is obviously *not* the kind of information which we understand with our conscious mind. Mental phenomena and our cognitive capacities, here, and the neural mechanisms in the brain,

there, do not share any obvious dynamic properties, for which a kind of a part-whole relation may be established.

Brain and mind, or our neurons and our ideas, do not stand in any known kind part-whole relation. Both are localized in our heads, but no spatial, dynamic, or causal relations between them are known, it is only possible to find and investigate specific correlations between them. So, cognitive neuroscience investigates the correlations between the neural activities in certain brain areas, on the one hand, and the contents of a test person's consciousness or certain cognitive capacities of a human being or an animal, on the other, and this is not in vain. In his book *Mental Mechanisms*, William Bechtel correspondingly emphasizes that cognitive neuroscience may employ heuristic identity assumptions about mental phenomena and their physical basis in neural mechanisms:

One of the virtues of viewing identity as a heuristic claim is that it can guide not only the elaboration of the two perspectives which are linked by the identity claim, but it can use each to revise the other. (Bechtel, 2008, 71).

This heuristic identity claim gives rise to the term “mental mechanism”. The respective heuristics is most fruitful for cognitive neuroscience, but to talk of “mental mechanisms” is here not associated with any mechanistic explanation proper discussed in this paper or in the recent debate.

10.5 Some Important Caveats

Yet it remains unclear how far we may go in generalizing genuine mechanistic explanations that indeed *explain* their explanandum from the causal components of a complex system. Moreover, I must confess that in the end it also remains unclear what a *genuine* mechanistic explanation is. In Sect. 10.3.3 I proposed to generalize mechanistic explanations in terms of the physical dynamics of a complex system and the conserved quantities of its causal components. However, to what extent can this explanation be generalized to higher-level mechanisms *beyond* physics? There are at least two more crucial caveats. Both are closely related to the limits of theoretical reduction.

First, in chemistry, biochemistry, and molecular biology, structural considerations become quite important for understanding mechanisms, and with them again the spatial structure of the causal components, which a physical dynamics neglects. In this sense, for higher level mechanisms the old concept of a mechanism as a machine is not completely off the mark, and so it is not completely metaphorical to speak of chemical, biochemical, or biological machines.

Second, neglecting the environment of a mechanism often leads to inadequate idealizations. This point becomes already evident in physics if we look at the decoherence approach to the quantum measurement problem (Bacciagaluppi, 2020). To speak of the mechanism of decoherence is to explain quantum measurements in (probabilistic) terms of the interaction between an entangled quantum

system-plus-measurement device and its environment. Understanding mechanisms often means looking not only top-down at the causal parts of a mechanism and their interactions, but also bottom-up at the way the mechanism is embedded or situated in its environment, as examples from higher-level sciences show, too (Bechtel, 2009; Bechtel & Abrahamsen, 2009).

10.6 Summary and Conclusions

The concept of a mechanism and the corresponding account of mechanistic explanations draw on the old analogy between machines and processes in nature. In view of scientific and technological progress, it is justified to generalize them from the traditional mechanistic explanations based on classical mechanics to current scientific practice. These generalizations have their counterpart in a generalized mechanistic methodology, which is typical of the “dissecting” sciences (Schurz, 2014, 35). This methodology, which aims first at decomposing natural phenomena *top-down* into lower-level causal components, and then, at giving *bottom-up* mechanistic explanations, traces back to the analytic-synthetic methods of early modern science, with Newton’s methodology as one of its most important roots. The method of dissecting the phenomena to explain them in mechanical terms became most successful in eighteenth and nineteenth century science. In twentieth century physics, however, the quantum revolution dispensed with the restriction of scientific explanations to classical mechanisms. Quantum mechanics provided new, generalized, mechanistic explanations, with quantum particles and field quanta as the causal components of mechanisms that explain the constitution of matter in terms of dynamic part-whole relations. These part-whole relations connect the properties of complex systems with the conserved dynamic quantities of subatomic particles. These conserved quantities satisfy well-defined conservation laws and support the definition of a mechanism in terms of sum rules that hold for conserved quantities. This definition generalizes Salmon’s account of mechanistic explanation to compound systems, and it specifies the definitions given by the proponents of the “new mechanisms” in terms of a physical dynamics.

This definition of a mechanism in a generalized sense is in accordance with the practice of atomic, nuclear, and particle physics, and it explains why and to what extent ontological reduction in physics is justified. But quantum fields with unsharp occupation number, continuous systems, and collective behavior such as phase transitions are beyond its scope, and it remains unclear whether it is more than a mere *façon de parler* to talk of the *mechanisms* underlying such phenomena. For the higher-level mechanisms of chemistry, biochemistry, and biology, further crucial limits of the approach must be considered, given the limits of theoretical reduction.

Another limit of mechanistic explanations not only for the approach suggested here, but also in a more general sense concerns the mental phenomena investigated by cognitive neuroscience. To talk of neural mechanisms indeed fits in with the

definition in terms of a physical dynamics. The underlying models are based on the laws of electrodynamics and electro-chemistry, and they are associated with the conserved quantities of charge and energy. To extend this talk to the relation between brain and mind, however, seems to be beyond the scope of any mechanistic explanation, as far as such an explanation seems to require that the *phenomenon explanandum* of a complex system and its causal components share at least *some* dynamic property.

Acknowledgements This paper is based on my previous work on scientific explanation in neuroscience (Falkenburg, 2012) and mechanistic explanations in physics (Falkenburg, 2019). I would like to thank Davide Vecchi for his critical comments of a previous version of my paper.

Literature

- Achinstein, P. (1991). *Particles and Waves. Historical Essays in the Philosophy of Science*. Oxford: Oxford University Press.
- Anderson, P. W. (1972). More is different. *Science, New Series*, 177, 393–396.
- Anderson, P. W. (1997). *Concepts in solids*. World Scientific.
- Bacciagaluppi, G. (2020). The role of decoherence in quantum mechanics. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 ed.) <https://plato.stanford.edu/archives/fall2020/entries/qm-decoherence/>
- Beaney, M. (2021). Analysis. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2021 ed.) <https://plato.stanford.edu/archives/sum2021/entries/analysis/>
- Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Psychology Press.
- Bechtel, W. (2009). Looking down, around, and up: Mechanistic explanation in psychology. *Philosophical Psychology*, 22, 543–564.
- Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 421–441.
- Bechtel, W., & Abrahamsen, A. (2009). Decomposing, recomposing, and situating circadian mechanisms: Three tasks in developing mechanistic explanations. In H. Leitgeb & A. Hieke (Eds.), *Reduction and elimination in philosophy of mind and philosophy of neuroscience* (pp. 178–190). Ontos.
- Collins, W. (2012). mechanism. In *Collins English dictionary - complete & unabridged* (10th ed.). HarperCollins Publishers. <http://www.dictionary.com/browse/mechanism>. Accessed 14 May 2017.
- Cotes, R. (1713). Editor's Preface to the Second Edition. In: Newton 1726, 385–399.
- Craver, C. F. (2007). *Explaining the brain. Mechanisms and the mosaic unity of neuroscience*. Clarendon Press.
- Crombie, A. C. (1953). *Robert Grosseteste and the origins of experimental science 1100–1700*. Clarendon Press.
- Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. MIT Press.
- Dowe, P. (2008). Causal processes. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Fall 2008 ed.) <http://plato.stanford.edu/archives/fall2008/entries/causation-process/>
- Engfer, H. J. (1982). *Philosophie als analysis*. Fromann-Holzboog.
- Falkenburg, B. (2007). *Particle metaphysics. A critical account of subatomic reality*. Springer.
- Falkenburg, B. (2012). *Mythos Determinismus. Wieviel erklärt uns die Hirnforschung?* Springer.

- Falkenburg, B. (2015). How do quasi-particles exist? In B. Falkenburg & M. Morrison (Eds.), *Why more is different. Philosophical issues in condensed matter physics and complex systems* (pp. 227–250). Springer.
- Falkenburg, B. (2019). Mechanistic explanations generalized: How far can we go? In B. Falkenburg & G. Schiemann (Eds.), *Mechanistic explanations in physics and beyond* (pp. 65–90). Springer Nature Switzerland.
- Geyer, C. (Ed.). (2004). *Hirnforschung und Willensfreiheit*. Suhrkamp.
- Glennan, S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, 44, 49–71.
- Hintikka, J., & Remes, U. (1974). *The method of analysis: Its geometrical origin and its general significance* (Boston studies XXV). Reidel.
- Hobbes, T. (1655). *De corpore*. In J. C. Gaskin (Ed.), *Thomas Hobbes: The elements of law, natural and politic; Ch. IV* (pp. 193–212). Oxford University Press. 1999.
- Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117, 500–544.
- Kaplan, D. M. (2018). Mechanisms and dynamical systems. In S. Glennan & P. Illaris (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 267–280). Routledge.
- Lindberg, D. C. (1971). Lines of influence in thirteenth-century optics: Bacon, Witelo, and Pecham. *Speculum*, 46, 66–83.
- Lindberg, D. C. (1976). *Theories of vision from al-Kindi to Kepler*. University of Chicago Press.
- Loose, J. (1993). *A historical introduction to the philosophy of science* (3rd ed.). Oxford University Press.
- Lucretius. (1570). *De rerum natura*, ed. by D. Lambin, Paris.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Malisoff, W. M. (1940). Physics: The decline of mechanism. *Philosophy of Science*, 7(1940), 400–414.
- Newton, I. (1726). *The principia. Mathematical principles of natural philosophy*. A new translation by I. Bernhard Cohen and Anne Whitman. Univ. of California Press. 1999.
- Newton, I. (1729). *Principia. Sir Isaac Newton's mathematical principles of natural philosophy and his system of the world. Vol. I: The motion of bodies. Vol. II: The system of the world*. Mott's translation revised by Cajori. Univ. of California Press 1934, 1962.
- Newton, I. (1730). *Opticks or treatise of the reflections, refractions, inflections & colours of light*. Based on the fourth edition: London 1730. Dover 1952, 1979.
- Odenbaugh, J., & Griffiths, P. (2022). Philosophy of biology. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2022 ed.). <https://plato.stanford.edu/archives/sum2022/entries/biology-philosophy/>
- Pappus of Alexandria. (1589). *Pappi Alexandrini Mathematicae Collectiones*. English edition: Pappus of Alexandria, *Book 7 of the Collection*. Part 1. Edited with Translation and Commentary by Alexander Jones. New York 1986. – Quoted after the translation in Hintikka & Remes 1974.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
- Salmon, W. (1997). Causality and explanation: A reply to two critiques. *Philosophy of Science*, 64, 461–477.
- Schiemann, G. (2019). Old and new mechanistic ontologies. In B. Falkenburg & G. Schiemann (Eds.), *Mechanistic explanations in physics and beyond* (pp. 33–46). Springer Nature Switzerland.
- Schurz, G. (2014). *Philosophy of science: A unified approach*. Routledge.
- Singer, W. (2003). *Ein neues Menschenbild? Gespräche über Hirnforschung*. Suhrkamp.
- Singer, W. (2004). *Verschaltungen legen uns fest. Wir sollten aufhören von Freiheit zu sprechen*. In Geyer 2004, 30–65.
- Roth, G. (2003). *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten steuert*. Suhrkamp.
- Rubia, F. L. (2009). *El fantasma de la libertad. Datos de la revolución neurocientífica*. Editorial Crítica.

- Unguru, S. (1974). Pappus in the thirteenth century in the Latin west. *Archive for History of Exact Science*, 13, 307–324.
- Vesalius, A. (1543). *De humani corporis fabrica libri septem*. Johannes Oporinus.
- Wigner, E. P. (1939). On Unitary Representations of the Inhomogeneous Lorentz Group. *Annals of Mathematics*, 40, 149–204.
- Worrall, J. (2000). The scope, limits, and distinctiveness of the method of ‘deduction from the phenomena’: some lessons from Newton’s ‘demonstrations’ in optics. *Brit J Phil Sci*, 51, 45–80.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 11

The Mechanisms of Emergence



Stuart Glennan

Abstract Emergentism is often imagined to be opposed to mechanism. If some phenomenon admits of mechanistic explanation, it is thought to be *ipso facto* not emergent. In this paper I argue to the contrary that emergence requires mechanism. Whenever some emergent phenomenon occurs, there is a mechanism responsible for its emergence. To make this case I show how mechanisms can explain four commonly held characteristics of emergent phenomena – dependence, autonomy, novelty and holism. By looking at the various kinds of emergence-generating mechanisms, it will be possible to classify different kinds of emergent phenomena by the particular features of the mechanisms that generate them, and so to bring some order to diversity of phenomena that we call emergent.

Keywords Emergence · New mechanism · Dependence · Autonomy · Novelty · Holism · William Wimsatt

11.1 Introduction: Mechanisms and Emergence

While there is no consensus on what emergence is or where and when it occurs, there is widespread agreement that it is opposed to mechanism. That opposition dates back to the foundations of the contemporary emergence debate in C.D. Broad's *Mind and its Place in Nature* (1925). Relatedly, whatever emergence is, it is generally understood to be opposed to reductionism. A reductive explanation of some phenomenon is taken to show that the phenomenon is not emergent.

In this paper I argue that these assumptions about emergence are misguided. Any emergent phenomenon must emerge *from somewhere*, and this somewhere is the mechanism that is responsible for the emergent phenomenon. I shall argue for this conclusion by showing how mechanisms can be responsible for phenomena that

S. Glennan (✉)
Butler University, Indianapolis, IN, USA
e-mail: sglennan@butler.edu

exemplify widely agreed upon criteria of emergence. Moreover, different kinds of mechanisms will satisfy these criteria in different ways and to different degrees, suggesting we can use kinds of emergence-generating mechanisms to distinguish different kinds and degrees of emergence.

The account I will offer here owes much to the work of Bill Wimsatt. Wimsatt's basic formulation of emergence is this:

Emergence of a system property relative to the properties of the parts of that system indicates its dependence on their mode of organization. It thus presupposes the system's decomposition into parts and their properties, and its dependence is explicated via a mechanistic explanation (Wimsatt, 2007, 276).

Wimsatt calls these emergent system properties “non-aggregative” because their dependence on mode of organization implies that you cannot aggregate the parts any which way and still recover the emergent properties. Purely aggregative properties – things like mass and charge – are rare, so emergence is the exception rather than the rule. Furthermore, Wimsatt sees no tension between emergence and reduction. For Wimsatt, *[a] reductive explanation of a behavior or a property of a system is one that shows it to be mechanistically explicable in terms of the properties of and interactions among the parts of the system* (ibid, 275; italics in original).

To defend a view like Wimsatt's, we must show how, despite the conventional wisdom about the opposition between emergence and mechanism, mechanisms can in fact explain how phenomena emerge. My strategy will be to identify four widely agreed upon principles for what is required for some phenomenon to be emergent, and then show how mechanisms can account for them.

The most commonly cited principles generally go by the names of dependence and autonomy (Gibb et al., 2019; O'Connor, 2021; Wilson, 2021). To these I add two additional criteria, that I call novelty and holism. These criteria are, on many taxonomies, aspects of the autonomy principle, but novelty, holism and autonomy can split apart and are sometimes in tension, so I will, following (Humphreys, 2016), keep all four. Here are one-sentence characterizations of each:

- **Dependence** — Emergence is a dependence relation between the source of emergence (the emergence base) and the result of the emergence (the emergent phenomena).
- **Autonomy** — Emergent phenomena should be autonomous from their emergence base.
- **Novelty** — Emergent phenomena have novel features that do not belong to the base from which they emerge.
- **Holism** — In emergent phenomena, the whole is more than the sum of the parts.¹

Here is the plan for the rest of the paper. In Sect. 11.2 I will provide a quick review of key features of mechanisms, as they have been articulated in recent discussions

¹Humphreys describes his four criteria as follows: “Emergent features result from something else, they possess a certain kind of novelty with respect to the features from which they develop, they are autonomous from the features from which they develop, and they exhibit a form of holism” (Humphreys, 2016, 26). I take his first criteria as expressing what I am calling “dependence.”

of mechanisms in the philosophy of science. I will use this account of mechanisms in Sect. 11.3 to interpret the dependence relation required for emergence as the relation of mechanism-dependence. I will clarify what can emerge from what by describing possible relata of this dependence relation, and I will use a basic distinction between two kinds of mechanisms to explicate the much-discussed distinction between synchronic and diachronic emergence. In Sect. 11.4, I will show how different kinds of mechanisms can generate phenomena that are in various ways autonomous, novel and holistic; I will also show how these three principles are related to some other features commonly held to be characteristic of emergent phenomena – for instance, downward causation, self-organization, and multiple realization. In the last section I will recap to what for some may be a nagging question – whether a mechanistic and reductionist theory of emergence misses the essence of emergence.

11.2 Mechanisms and their Varieties

The basic supposition of a mechanistic theory of emergence is that emergent phenomena emerge out of the activities of mechanisms, and that different varieties of emergent phenomena can be identified by comparing the kinds of mechanisms which generate them. In order to make a case for such a theory, I will begin with a review the basic features of mechanisms as articulated within the new mechanist literature in philosophy of science (Craver & Tabery, 2016; Glennan & Illari, 2018a; Glennan et al., 2021) and to describe briefly an approach to classifying varieties of mechanisms that I have developed elsewhere (Glennan, 2017).

In ordinary English usage, the word ‘mechanism’ has two senses. In the first, mechanisms are systems with interacting parts, often but not always artifacts; in the second, mechanisms are processes which bring about some happening or activity. Example of mechanisms in the first sense are things like clocks or computers; examples of mechanisms in the second sense are mechanisms of protein synthesis or reproductive mechanisms. Call these two senses respectively the systemic and the processual sense of mechanism.²

Within most scientific contexts, the processual sense of mechanism is the more common, and for this reason, new mechanistic approaches have taken the processual sense to be primary. The new mechanistic account of mechanisms can be briefly summarized in a definition I call minimal mechanism: “a mechanism for a

²The Oxford English dictionary lists these two definitions for mechanism:

1. a system of parts working together in a machine; a piece of machinery
2. a natural or established process by which something takes place or is brought about

These definitions suggest that the system sense is connected to machines and machinery, while the processual sense is connected with natural phenomena. While there is no doubt that many of our stereotypic examples of mechanistic systems are machines, there is nothing in the idea of a system that requires it to be of artificial construction; and similarly, there is nothing in the idea of a mechanistic process that requires it to be “natural.”

phenomenon consists of entities ... whose activities and interactions are organized so as to be responsible for the phenomenon” (Glennan, 2017; Glennan & Illari, 2018a)

According to minimal mechanism, mechanisms are individuated by what they do — their phenomena or behavior. One speaks of the mechanism of protein synthesis, or predation or reproductive mechanisms, or the mechanisms by which animals communicate, or by which national banks control the money supply. All of these are phenomena that depend upon mechanisms.

The mechanisms responsible for such phenomena are made up of constituents that are termed **entities**, **activities**, and **interactions**. Entities are understood to be “things” — objects, systems, structures, etc.³ They could be proteins, cells, organisms, families, baseballs, televisions, planets, stars or galaxies. The activities and interactions are the doings in which these entities engage — e.g., bonding, folding, striking, heating, walking, eating, radiating, exploding. Activities are processual in the sense that they are temporally extended, often but not always with distinct beginnings, intermediate stages and ends.

The difference between activities and interactions is simply the number of actors. Interaction requires multiple entities to be involved, whereas activities may involve only a single entity; for instance, sexual reproduction takes two actors, while asexual reproduction takes only one. Going forward, I will sometimes speak generically of activities as being inclusive of both one-place “solo” activities and multi-place interactions.

In order for a mechanism to give rise to some phenomena, its constituent entities, activities and interactions must be organized in a particular way. When, for instance, an animal turns its body, this activity (the mechanism’s phenomenon) requires that the parts of the animal (its joints, limbs, muscles, etc.) are put together in a certain way, and that the timing of the activities of and interactions between these parts are coordinated. The general lesson is that a pile of mechanism parts does not make a mechanism. This is the minimal sense in which mechanisms are always more than the sum of their parts.

It is helpful to think about mechanisms as composite processes. They are processes, because mechanistic phenomena always involve activities and interactions, which are temporally extended doings. They are composites, because these doings depend upon a mechanism constituted by its components — organized activities and interactions of some set of underlying entities; these components are said to be **constitutively relevant** to the mechanism.

Take, for instance, protein synthesis. This is an activity or process, and it is constituted by the activities and interactions (transcription, translation, etc.) of the various entities (DNA, mRNA, etc.) that are collectively responsible for the activity of

³This use of the term ‘entity’, which originates with (Machamer et al., 2000), is unfortunate given the fact that metaphysicians often use ‘entity’ to refer generically to anything that belongs in one’s ontology — including objects, but also properties, relations, laws, tropes, and whatever. I will occasionally switch between the new mechanist’s and the metaphysician’s sense, trusting to the context to make the intended meaning clear.

protein synthesis. It is because mechanistic phenomena always involve activity that Craver et al. (2021) suggest that all mechanisms have a “processual core.”

Entities are also composites, but they are not mechanisms in the processual sense of minimal mechanism.⁴ Their components are other entities. For instance, a cell is an entity composed of its membranes, organelles, and so on. These composite entities are generally what I’ve called **mechanistic systems** — composite entities whose persistence and interactions with the world depend upon mechanisms constituted by the activities and interactions of their components. A cell, for instance, is a mechanistic system because it cannot live and perform its functions within its environment without the constant operation of mechanisms involving organized activities and interactions of its parts.

One of the chief payoffs of a mechanistic account of emergence is that it will allow us to classify different varieties of emergence as arising from different varieties of mechanisms. The terms within the minimal mechanism definition suggests four dimensions of classification (Glennan & Illari, 2018b):

- the kinds of phenomena for which the mechanism is responsible
- the kinds of entities the mechanism has as constituents
- the kinds of activities and interactions their constituent entities engage in
- the ways in which entities and activities/interactions are organized within the mechanism

A fifth dimension concerns not current features but history. Mechanisms may be classified etiologically, by how they came to be.

These five dimensions are largely independent, so, for instance, it is possible for mechanisms with very different kinds of constituent entities and activities/interactions to have similar kinds of organization. In what follows we shall see that we can understand different varieties of emergence chiefly in terms of different kinds of phenomena, different kinds of organization and different kinds of etiology.

11.3 Emergence as Mechanism-Dependence

Minimal mechanism holds that all mechanisms are mechanisms **for** some phenomenon, and any such phenomena can be said to be **mechanism-dependent**. A mechanistic account of emergence proceeds from the supposition that the dependence between emergent phenomena and their emergence bases are relations of mechanism-dependence. Mechanism-dependence is not sufficient for emergence, since by itself it does not guarantee autonomy, holism or novelty, but it is necessary. In this section I will show how recognizing different varieties of mechanism-dependence relations and different possible relations yields a natural way of describing different varieties of emergence.

⁴See (Glennan, 2021) for my preferred account of this composition relation.

11.3.1 Producing versus Underlying and the Distinction between Diachronic and Synchronic Emergence

Perhaps the most basic distinction to make between kinds of mechanisms is the distinction between mechanisms that produce phenomena and mechanisms that underlie phenomena. In the former case, there is a mechanistic process ψ that is triggered by some set of startup conditions ψ_{in} and terminates with a product ψ_{out} .⁵ For instance, in a protein synthesis mechanism, ψ_{in} would be the set of conditions initiating protein synthesis, and ψ_{out} would be the protein product. In the latter case, the phenomenon in question is the activity or process itself, and the mechanism underlying it is the set of organized activities and interactions of entities that constitute that activity. For instance, if a muscle (S) contracts (ψ) then what underlies this contraction (i.e., what constitutes it) are the contractions (ϕ_i) of the many muscle fibers (X_i) that make up the contracting muscle.

The distinction between producing and underlying relations corresponds to the familiar distinction between diachronic and synchronic emergence (see e.g., Humphreys, 2016, sec. 1.7.4). In diachronic emergence, we can interpret the emergence base as the set of startup conditions which, via an etiological mechanism, produces the emergent phenomenon. The emergence base is temporally prior to and distinct from the emergent phenomenon, and the etiological mechanism is the causal process by which the emergent phenomenon emerges. In synchronic emergence by contrast, the emergent phenomenon depends upon an underlying mechanism, which coexists with the phenomenon in space and time. This interpretation of synchronic emergence permits the relation between the emergence base and the emergent to be temporally extended and dynamic. If, for instance, a behavior of an animal emerges synchronically from the activities of the animal's parts, the behavior and the underlying mechanism will both involve temporally extended activities and interactions.

The same phenomena may emerge both diachronically from a temporally prior emergence base and synchronically from a temporally overlapping emergence base. An organism is a mechanistic system that emerges in both of these senses; it emerges diachronically from developmental mechanisms and synchronically from the entities and activities which underlie and maintain the organism and its activities. The two varieties of emergence correspond to the two aspects of mechanistic explanation. One explains diachronic emergence via an etiological mechanistic explanation, while one explains synchronic emergence via a constitutive mechanistic explanation.

⁵I make use here of Craver's (2007) formalism that denotes activities by Greek letters and entities by Roman letters. Since many mechanistic processes are cyclical (as for instance in metabolic pathways), the startup and termination points may be arbitrarily identified points within ongoing cycles.

11.3.2 *What Emerges: The Relata of Mechanism-Dependence Relations*

Since emergence is a kind of dependence relation, one way to distinguish its varieties is by the varieties of relata that may stand in this relation. In the metaphysics literature the relata are often assumed to be properties, but interpreting emergence as mechanism-dependence suggests that things other than properties can stand in emergence relations. We can sort emergent relata into the following categories:

- Emergent processes, activities and interactions
- Emergent entities or systems
- Emergent properties and relations

Each of these kinds of emergents may emerge both synchronically from an underlying mechanism or diachronically from an etiological mechanism.

Consider first activities. A mouse running through a maze is a standard example of what in the mechanisms literature is called an “entity acting” (Krickel, 2018). The mouse’s running (S ψ -ing) depends upon the orchestrated activities and interactions of the entities that underlie this ψ -ing. These include activities of and interactions between elements of the muscular/skeletal system, the cardiovascular system and the central nervous system, among others. The relation between the mouse’s running and the activities and interactions of its parts is synchronic and constitutive. When the mouse moves a leg, that movement is not the cause of the mouse’s running but is part of the activity of its running.

While some activities are aptly described as entities acting, not all are. Many activities are in fact interactions between two or more entities (in which case we would say, e.g., that S and T are ψ -ing together). One example is the mechanism by which one or more neurons trigger another neuron across a synapse. In such a case there are constituent entities within S and T, as well as entities in the environment (e.g., neurotransmitters in the synapse) whose activities and interactions underlie the interaction between the triggering and triggered neurons.

The mouse running or the neuron triggering are plausible candidates for synchronically emergent activities, though to make the case fully we would have to consider how they might meet the autonomy, novelty and holism requirements which we will take up in the next section. It is less plausible to think that particular instances of such activities emerge diachronically. We give etiological mechanistic explanations of these processes, but they would largely be accounts of the events that produce the startup conditions for these mechanisms. For instance, we could give an etiological explanation that identified the stimulus that made the mouse start running. But such an explanation is straightforwardly causal, and something more than bare causal dependence is needed for emergence.

Better candidates for diachronic emergence of activities are atmospheric processes like winds, rains and hurricanes. Consider as an example the trade winds. The trade winds are relatively constant easterly winds that blow in the tropical

regions north and south of the equator. They are generated by an underlying mechanism involving both the cycling of air and moisture from the equator towards the poles and the earth's rotation. The conditions that give rise to these winds have been fairly constant throughout recorded human history, but they depend for their existence upon features of the planet and atmosphere that have changed over time. For instance, the current wind patterns depend upon the location of continental land masses, and those have shifted over time. Because these historical conditions generated processes that are novel and self-organizing, the resulting trade winds seem like plausible candidates for diachronically emergent processes.

The second category of emergents is entities/systems. In what sense might they depend upon mechanisms, and thereby emerge from the activities of mechanisms? Most obviously, entities can be the products of producing mechanisms: proteins are produced by a synthesis mechanism, cars are produced by an assembly line, and organisms are produced by reproductive and developmental processes. If these entities are indeed emergents, this variety of emergence would be diachronic. Such products depend diachronically upon their antecedents, and they "emerge" in some pre-theoretical sense as the product of the mechanism. Which such products should count as genuinely emergent will depend upon the degrees and respects in which the products meet the autonomy, novelty and holism criteria.

It is less obvious whether and when entities can rightly be said to emerge synchronically from constitutive mechanisms, but if the entity is what I've called above a mechanistic system, it is plausible to say that the system emerges from the activities of its constituents. Consider a mouse. While the mouse is not a mechanism in the processual sense of minimal mechanism, its persistence as a living mouse requires the action of many mechanisms involving its constituent entities and their actions and interactions, both among themselves and with their environment. A mouse is sustained for instance by its metabolism, by its moving about to find food and evade predators, and so forth. Since its continuation as a living mouse depends synchronically upon these mechanisms, there is a clear sense in which we can understand these mechanisms as an emergence base from which the living mouse synchronically emerges.

What of emergent properties and relations? Certainly there are such things, but a mechanistic theory of emergence gives a different account of their nature than most metaphysical accounts. Whereas abstract metaphysical accounts of emergence conceive of emergence relations as modal relations like supervenience or grounding, the mechanistic account understands properties to belong to composite entities. It takes a broadly causal/dispositional view of properties — a view that properties are individuated by their causal role. To say, for instance, that the table is solid, is to say that other objects will not fall through it; or to say that a piece of paper is flammable is to say that it will catch fire when touched by a flame. The mechanist's insight is simply to say that composite entities have such dispositions in virtue of being mechanistic systems — systems within which mechanistic processes involving their components can be triggered to manifest such dispositions. A paper is flammable because interactions like touching it with a lighted match will trigger its burning,

while a fireproof cinder block, because it is composed of different kinds of entities, differently organized, will react differently to a flame. If this view of properties is correct, then all but the most fundamental properties will be properties of composites, and their manifestation will depend upon mechanisms. Relations between properties and the mechanisms upon which they depend are synchronic, because properties manifest themselves through the activity of underlying mechanisms.

Properties also emerge diachronically from etiological mechanisms that produce changes in the properties of entities. Consider again the mouse and its running. The mouse's capacity to run is a behavioral property of the mouse — a disposition that manifests itself when some stimulus or cognition triggers it. This disposition is not something the mouse is born with; it emerges gradually as the mouse develops and interacts with its environment. More generally, when we consider the traits of biological organisms (anatomical, physiological, or behavioral), we will find that they emerge in individual organisms via developmental mechanisms of various kinds. A different kind of diachronic emergence occurs with respect to species. On evolutionary time scales, lineages of organisms acquire novel traits via evolutionary mechanisms — most commonly the mechanism of natural selection.

Diachronic emergence of properties is not necessarily limited to the biological. Many physical and social entities or systems acquire novel properties through the diachronic operation of mechanistic processes — mountain ranges, atmospheres, stars, galaxies, political parties, economic markets, and nation-states, to name just a few. Of course, the mere fact that some entity acquires properties gradually via a mechanistic process is not sufficient to classify it as emergent, but these examples at least suggest the possibility of the diachronic emergence of properties.

One last category of emergent that is important in many scientific discussions of emergence is the emergence of patterns (Winning & Bechtel, 2019). Patterns are of a higher order than entities, activities or properties, because when a pattern emerges, it can be a pattern of any of these things. For instance, there can be a pattern in distribution of any of these things. Take for example oscillatory patterns. When pendulums oscillate, what oscillates is the position and momentum of the weight; when circuits oscillate, what oscillates is current; and when populations oscillate (say due to procreation and predation), what oscillates is the size or density of a population within its environment; when markets oscillate, what oscillates are supply, demand and price. Other examples of emergent patterns include patterns of motion in flocks of birds or schools of fish, or, to consider a purely digital example, patterns in the emergence, disappearance or motion of shapes within cellular automata, exemplified by Conway's game of life.

Because it is higher order, the tools for studying pattern emergence are abstract and mathematical, including dynamical systems theory, chaos theory, cellular automata, and network analysis. These mathematical theories provide domain-independent tools for describing patterns in and dynamics of systems, processes, events, etc. They are especially suited to describing cases of diachronic emergence of patterns, like transitions between stable and chaotic behavior, or phase transitions

in states of matter. From a mechanistic point of view, patterns emerge because of how mechanisms are organized, and these mathematical tools allow one to describe these patterns in an abstract and general way.⁶

11.4 Autonomy, Holism, and Novelty in Mechanistic Emergence

In the previous section I've argued that the dependence required for emergence should be understood as mechanism-dependence. In this section I turn to the other three not wholly independent criteria – autonomy, holism and novelty. My strategy will be to run through a series of suggestions about what features might be required in order for a system to meet these criteria, and show how they can be fruitfully explicated within the mechanistic framework.

11.4.1 *Non-aggregativity*

Consider first Wimsatt's suggestion (1997, 2000, 2007) that emergence occurs in systems whenever they are **non-aggregative**. A property of a system is non-aggregative to the extent that its existence or value depends upon how the parts of the system are arranged. Wimsatt interprets aggregativity as a kind of invariance or stability of properties under various kinds of manipulations or transformations of a system's parts. He identifies conditions required for aggregativity:

1. *IS (InterSubstitution)*: Invariance of the system property under operations rearranging the parts in the system or interchanging any number of parts with a corresponding numbers of parts from a relevant equivalence class of parts ...
2. *QS (Size Scaling)*: Qualitative similarity of the system property (identity, or if a quantitative property, differing only in value) under addition or subtraction of parts....
3. *RA (Decomposition and ReAggregation)*: Invariance of the system property under operations involving decomposition and reaggregation of parts
4. *CI (Linearity)*: There are no Cooperative or Inhibitory interactions among the parts of the system that affect this property. (Wimsatt, 2007, 280–281)

⁶Mechanistic approaches are sometimes seen as at odds with more abstract and mathematical approaches to pattern emergence, especially with regard to explanation. Explanation of emergent phenomena is often seen to depend upon getting away from mechanistic detail (O'Malley & Dupré, 2005; Silberstein & Chemero, 2013; Batterman & Rice, 2014), and disciplines like systems biology and systems neuroscience are seen as alternatives to mechanistic approaches. While I cannot argue the case here, it seems to me that this tension is much overblown, and in fact that tools such as dynamical systems theory are an important part of the toolkit for describing mechanisms of emergence. For further discussion see (Kaplan & Bechtel, 2011; Craver & Kaplan, 2018; Brigandt et al., 2018)

Any failure to meet these conditions would imply emergence, with different failures yielding different varieties of emergence. On this view, emergence is ubiquitous, and non-emergent properties are the exception rather than the rule, since most properties of composite systems depend not just upon what parts the composite has, but how those parts are arranged.

A homely example should make clear why non-aggregativity is the usual case. Lawn mowers have many properties, but perhaps the most salient is that lawnmowers can cut lawns. But the ability of the lawnmower to cut lawns depends not just upon the parts which make up the lawnmower, but upon those parts being assembled in the right way. Lawn-mowing is an activity which the lawn mower can engage in but which its components cannot. Moreover, you cannot typically take away parts of a lawnmower in a way that gradually degrades its lawn-mowing capacity.

Although Wimsatt claims all non-aggregative properties are emergent, it isn't clear that non-aggregativity alone guarantees that a system will meet all of the four conditions identified at the outset. Clearly it meets the dependence criterion, since emergent properties mechanistically depend upon the components of systems and mechanisms. Equally clearly, non-aggregativity is a kind of holism. If "aggregation" is analogous to summing, non-aggregativity entails that the whole is more than the sum of its parts.

Novelty, however, does not seem guaranteed by non-aggregativity. It is true that organized mechanical systems like the lawn mower will have properties and abilities that their parts do not – only the whole lawnmower can mow the lawn – but some failures of aggregativity don't yield genuinely new properties. For instance, in an electrical circuit, changing arrangements of resistors from serial to parallel changes the amount of current flowing, but it won't introduce a new kind of property. Most importantly, there is no obvious link between non-aggregativity and autonomy. The fact that properties of a system are organization-dependent does not by itself seem to imply any sense in which the system is autonomous from the components upon which it depends.

We may conclude I think that emergent properties will necessarily be non-aggregative, but that non-aggregativity is not a sufficient condition for emergence. Alternatively one might take non-aggregativity to be a weak form of emergence

11.4.2 Externalism

Non-aggregativity implies a weak more-than-the-sum-of-its-parts kind of holism, but it certainly does not block reductive explanation. Perhaps other sources of holism will yield stronger varieties of emergence. One possible source much discussed in philosophy of mind and cognitive science is externalism. Externalism is generally taken to come in two varieties. Passive externalism is a view about mental content, exemplified by Putnam's (1975) motto that "meanings ain't in the head." More recent discussions have focused on "active externalism" — sometimes called theories of 4E (embodied, embedded, extended, enacted) cognition (Newen et al., 2018).

Active externalist theories point to the ways that cognitive agents solve problems or complete tasks using resources that extend beyond their brain, and often beyond their body. Examples include the ways humans use fingers or pencil and paper or calculators to help solve math problems, the way that our bodies rely on environmental affordances to simplify search and movement tasks, and the ways that perception of objects requires movement about and interaction with those objects. The common theme of externalist approaches to cognition is that they suggest that an account of cognitive processes cannot be given which localizes cognition within the “naked brain.” Interactions with entities outside of the brain is required for an agent to acquire or exhibit a cognitive capacity. This is a distinctive kind of holism, since we find a property (here a cognitive capacity) that we typically ascribe to a cognitive agent in fact depends upon the agent acting within an encompassing environment. 4E cognition is amenable to mechanistic explanation (Miłkowski et al., 2018), but the mechanisms are necessarily wider than those working within the brain alone. These emergent capacities are also novel in the straightforward sense that they only emerge as the brain interacts with its environment.

While active externalism is a cognitive phenomenon, it seems to be an instance of something that occurs in many systems, a phenomenon we might call emergence as non-locality. This kind of emergence occurs whenever some activity or capacity that is often attributed to an entity in fact depends for its existence upon features of or interactions with the world beyond the entity’s boundaries. For instance, if it is in fact that case that scientific knowledge is essentially social in character, it follows that this kind of knowledge can only arise through organized interactions of scientific communities, rather than being localized in individual scientists. Similarly, in ecology, resilience is a property that belongs to an ecosystem rather than its parts. While individual organisms or species may be resilient in certain respects, the ecosystem’s ability to recover from shocks and stresses cannot be reduced to the resilience of its parts.

11.4.3 Downward Causation

Another feature widely held to characterize systems with emergent properties is downward causation. Many common sense and scientific causal claims appear to express relations in which higher level activities, events or properties causally influence happenings at a lower level. Mental causation seems to have this character. Undertaking a meditative exercise can have effects on physiological processes like heart beats. More generally, any time a perception or decision is followed by bodily action, as when my fingers press the keyboard as I try to explain downward causation, mental events seem to produce physical changes in the body. Moreover, downward causation is by no means limited to the mental. In fluid dynamics, large scale convection currents seem to causally constrain elements within those currents (Bishop & Silberstein, 2019). Social facts or events seem to exert downward causation on individual persons (Sawyer, 2004). In ecology and evolution, system level properties like population density affect individual fitness (Millstein, 2006).

A significant source of the recent resurgence of interest in emergent phenomena has been the need to make sense of these kinds of causal relations. Much of it has been motivated by Kim's causal exclusion argument, which was developed as an objection to non-reductive physicalism (Kim, 1993). Kim claims that if mental or other higher-level properties supervene on physical properties, the mental or higher-level properties cannot have causal efficacy, since the physical properties upon which they supervene are already sufficient for their effects.

There is now a substantial philosophical literature which attempts to elucidate downward causation as it occurs in disciplines across the physical, life and social sciences, with much of it aiming to show that the phenomenon is ubiquitous, non-mysterious and explicable by looking at the structure of mechanisms (Ellis et al., 2012; Paolini Paoletti & Orilia, 2017). Craver and Bechtel (2007), for instance, argue that top-down causation is actually a hybrid between interlevel mechanistic constitution relations and intralevel causal relations which they call "mechanistically mediated effects." Others argue that in certain kinds of mechanisms, higher level system or process variables are the causally relevant or difference making features, and form the basis for explanation and intervention (Glennan, 2010; Woodward, 2021). Relatedly, others have tried to understand the notion of higher level cause in terms of concepts of constraint and control (Kistler, 2009; Bechtel, 2017).

While there are disagreements about the proper metaphysical interpretation of top-down causation, there is a broad consensus that top-down causation is a widespread phenomenon, and that its occurrence is explained by the mechanistic structure of systems and processes in which it occurs. Different varieties of top-down causation may arise as the result of different kinds of mechanisms (Ellis, 2011). However exactly we explicate it, downward causation seems to be an important mark of the emergent. Top-down causation implies a kind of holism (system level properties cause, control or constrain activities of the parts) and a kind of autonomy (system level properties are the properties that make a difference). Moreover, systems or processes that exhibit top-down causal influence acquire properties and capacities that are novel relative to the properties and activities of their constituents.

11.4.4 Self-Organization

Another commonly cited characteristic of emergent systems and processes is **self-organization**. The concept of self-organization has a long history (Keller, 2008) and is not easily defined, but the general feature of self-organizing systems (entities) and processes (activities) is that they form and maintain themselves in the absence of external control. These processes of formation and maintenance arise spontaneously out of the local activities and interactions of components of these systems and processes. Examples of self-organization occur across a range of domains. Some examples include processes of crystal formation, processes of membrane formation

in cells, the processes that direct collective movements of herds of animals or flocks of birds, and the economic processes that lead to the formulation of markets and the establishment of prices.

A precondition for systems and processes to self-organize is a feature of mechanistic organization I call “affinitive organization” (Glennan, 2017). An interaction is affinitive to the extent that it is directed by the dispositions of the interactors rather than the direction or arrangement of an external controller. Consider as an example the difference between a cellular membrane and a brick wall. The membrane self-organizes because of its component entities, phospholipid molecules have hydrophilic heads and hydrophobic tails, which will in an aqueous environment spontaneously aggregate into sheets. Bricks on the other hand are not moved by attractive and repulsive forces to line up; you need a bricklayer to line them up and cement them together. It is precisely this absence of a controller that yields the sense in which self-organizing systems and processes are autonomous.

A related kind of organization is **self-maintenance**. A self-maintaining system is one that carries with it the capacity to maintain its properties and functions, repair damage, and so forth. Organisms are paradigms of self-maintaining systems; they have the capacity to maintain their state (e.g., concentrations of metabolites, temperature), to repair damaged tissues, and to destroy infectious agents. In contrast to a car, you do not normally have to take your body to the shop to repair a scratch.

11.4.5 Multiple Realization and Dynamical Autonomy

Multiple realization arguments have been employed to defend non-reductive physicalism, and, to the extent that that emergence is understood to be opposed to reduction, multiply realizable properties are natural candidates for emergents. Realization is a dependence relation, and it is natural to interpret it as a species of mechanism-dependence. Consider a well-known toy example, the capacity of certain tools to remove corks from wine bottles (Shapiro, 2000). This capacity is realized by different kinds of mechanisms in different kinds of corkscrews. For instance, a waiter’s corkscrew operates by twisting a screw into the center of the cork, swinging a hook attached to the device’s handle onto the bottle’s lip, and pulling up on the handle, which uses the hook as the fulcrum of a lever to pull out the cork. A winged corkscrew by contrast has a body with a collar that can rest on the lip of the bottle, and a screw mounted in the center of the body that can move through the collar as it is twisted into the cork. When the screw is inserted, two levers attached to the screw by gears (the wings) are driven up, and pushing down on the wings pulls the cork out through the collar.

Multiple realization allows for token reductive explanations, since each token system (like a corkscrew) has a token mechanism that gives it that capacity. Multiple realization arguments are instead focused on types. A given type of capacity, like the capacity to remove corkscrews, can be realized by many different types of mechanisms. Historically the point of multiple realizability arguments was to argue for the

explanatory autonomy of special sciences (Fodor, 1974). Explanations in economics, for instance, can formulate principles about the relationship between supply, demand and price, without concerning itself with the physical realizations of money or the mechanisms by which money and goods trade hands.⁷

Plausibly, multiple realizability is enough to yield a kind of autonomy sufficient for emergence. While any token of the emergent capacity will depend upon a particular realizing mechanism, different tokens of this capacity can and often do depend upon different types of mechanisms. This means the capacity as such does not depend upon a particular mechanism, and is hence autonomous from it.

What Wimsatt calls dynamical autonomy is a conceptual cousin of multiple realizability, but while multiple realization focuses on how different tokens of a particular property or kind can be realized by distinct mechanisms, dynamical autonomy points instead to the ways in which a single token macro-state or property of a system can persist even in the face of changes to that token's micro-state and mechanisms. As Wimsatt puts it "dynamical autonomy ...entails that most ... micro-level changes don't make a causal difference at the macro-level (Wimsatt, 2007, 218).

Examples of dynamical autonomy abound. For instance, the states of an organism can remain stable at the macro level even as micro-level changes like the birth and death of cells are constantly occurring. Similarly, psychological states of human being can persist even as many neurological features of the person are in flux. It is this persistence that guarantees that it is the macro-states rather than the micro-states which are causally relevant to a system's behavior. Dynamical autonomy seems to capture the both the dependence and autonomy required for emergence. There is dependence, because the macro-state needs a micro-state to realize it, but there is autonomy, because the macro-state and its causal powers can persist in the face of changes to the micro-state.

11.4.6 Transformation and Fusion

Paul Humphreys argues that philosophical accounts of emergence have too often ignored diachronic emergence, and offers as a remedy an account of what he calls transformational emergence (2016, sec. 2.1.1). Humphreys motivates the account by considering the emergent behavior of mobs. What we observe in so-called "mob psychology" is that the behavior of individuals within the mob is transformed, so that those whose normal dispositions might be friendly and non-violent exhibit a

⁷Multiple realizability arguments have not aged that well, because it is increasingly apparent that the implementing mechanism makes a difference to the realized capacity (Polger & Shapiro, 2016). Not all corkscrew mechanisms are equally good at removing corks, and all have their quirks. Or, more seriously, it is evident that changes to the medium of financial transactions can profoundly affect the market's behavior. But even if multiple realizable properties are harder to come by than was once thought, it may be still that such properties are good candidates for emergents.

new and destructive set of dispositions when they become part of a mob. As Humphreys describes the case, what emerges is a transformed individual behaving in accordance with a new set of generalizations – laws of abnormal psychology rather than of ordinary psychology. After such a transition, the transformed individuals act in ways that are quite different from their past selves.

Humphreys goes to some length to distinguish the transformation that occurs in individuals in a mob from the change in behavior that one sees in ordinary crowds or in flocks of birds. At first sight they seem similar, since crowds and flocks, just like mobs, make their members behave in ways they would not on their own. But Humphreys claims that only in the case of the mob is there genuine ontological emergence. The reason he thinks is that in ordinary crowds or flocks nothing essential has changed in the dispositions of the individuals, while in the mob something has. Consideration of our own experience suggests that this distinction is important. If I am moved along in a crowd as I leave the stadium, my motion is severely constrained, so I have lost my individual agency, while, in contrast, were I caught up in the feelings of the mob, I might want to engage in the violent acts that others were engaging in; not just my actions but my beliefs and desires would be transformed.

Despite its initial plausibility, Humphreys' distinction between essential and accidental transformations is hard to sustain. The account presupposes that the individuals that form some composite have a set of essential and intrinsic characters that, together with their arrangement, fix the properties of the composite of which they are part. But this kind of essentialism is something we generally have reasons to reject. Whenever a composite is formed, it constrains and alters the components' activities, and it is extremely difficult to make a principled case that some but not all such transformations represent the transformation of the component itself. All we can say with certainty is that the component's behavior is transformed by its placement in the composite.⁸

Humphreys has identified a particularly strong form of transformation he calls fusion. He considers covalent bonding to be a paradigm case. Consider a hydrogen molecule consisting of two hydrogen atoms. Humphreys writes that when the atoms are close together, their individual identities disappear:

Although the standard treatments usually talk of two indistinguishable electrons, what one really has is a joint probability distribution within which there is no sense to be made of separate, coexisting particles. This means that if the original, spatially separate hydrogen atom 1 is identified with proton 1 and electron 1, and the original, spatially separate hydrogen atom 2 is identified with proton 2 and electron 2, within the bonded molecule it is no longer possible to say that hydrogen atom 1 (or 2) exists as an identifiable subunit. (Humphreys, 2016, p. 83).

⁸Although Humphreys (2016) uses mob psychology to introduce the notion of transformational emergence, he ends up doubting that it is a genuine case. His rationale is that an improved neuropsychological theory may show that the essential properties of individuals do not change in mobs, but only that they only manifest themselves in novel ways in mob-like environments. Given my skepticism that there are any essential psychological properties (and with them laws), I would argue that mob psychology is a good example of transformational emergence.

If this description is right, it does suggest that in this kind of bond, the identity of the components is lost in the fused entity. What emerges is a new entity — not just an arrangement of existing entities. Humphreys sees this a special case of transformation, where the individuals in the composite are not merely changed in their essential properties, but disappear into a new entity. Such processes, Humphreys plausibly believes, exemplify a strong sense of novelty and holism.

It is not obvious to me though that there is a principled and domain-neutral account as to when acts of composition lead to the disappearance of the components. Humphreys for instance argues (84–85) that when one kneads together two lumps of clay, the original lumps may not be lost, because at a molecular level the original lumps (now spatially distributed) could in principle be recovered. In contrast, speculatively, he suggests (86) that the unity government formed in the United Kingdom during World War II was an act of fusion, in the sense that the Labor and Conservative parties “disappeared” and were replaced by a new party with novel properties. My intuition is that properties of MPs might be more recoverable than properties of clay lumps, and that the properties of the unity government, as well as transformation of its members might be mechanistically explained. But whatever one’s intuitions, it seems that absent a firmer account of essential properties, all we will know is that in these transformational processes, both the components and the composites will change in dispositions and activities so as to exhibit the properties of novelty, autonomy and holism.

Humphreys’ examples of diachronic transformations are drawn primarily from the physical sciences, but the model is intended to be general, so it is worth mentioning a few much-discussed cases from other domains to which Humphreys’ account might be applied. One important case is intentionality. Just what intentionality is and where it comes from is of course a matter of controversy, but a prominent approach that seems to involve diachronic transformation is Dretske’s (2021) account of a “recipe for thought.” Dretske holds that a system acquires thoughts and other intentional states as it interacts with features of the environment which are relevant to its needs or interests. The system starts with some primitive capacities, but it only acquires genuine intentional states when environmental processes trigger these capacities. If the system has the capacity to rewire its responses based upon this input, it should over time begin to acquire something like genuine representations (and misrepresentations) — or so the theory goes.

Dretske’s recipe for thought is meant to account for how thought and representation develop in natural systems, but we can tell a similar story for artificial systems that are capable of being trained. The deep neural networks used in speech or facial recognition technologies are systems that use their environment (a set of training data) to tune the network to recognize or classify words or faces. Similar techniques are used by chat bots and game-playing AIs that learn from their experience as they interact with you. While it seems improbable that these kinds of systems yet have the kinds of goals and interests required to acquire “genuine” intentionality, they certainly acquire novel capabilities through their interactions with the environment. The reason that it is so natural to call these sorts of phenomena emergent is that one can’t really build in the capacity from the start. As Dretske emphases, you can’t just

add such a capacity “the way you add spices in a recipe for lasagna. Adding the function is more like *waiting* for the dough to rise” (2021, 357).

One last example worth mentioning is what Paul (2014) has called “transformative experiences.” Paul argues that humans cannot make rational decisions (in the sense of decision theory) regarding experiences – like having a child or undergoing a religious conversion– which will be both epistemically and personally transformative. Personally transformative experiences can make changes to one’s basic psychology – to one’s likes, values and personality – such that one emerges as a qualitatively different person. Following Humphreys’ model of transformation, such personal transformations, when they occur, would be cases of diachronic emergence. Paul’s account may run into the same trouble that Humphreys had with distinguishing genuinely transformative changes from mere changes in behavior, but, to the extent that such a distinction can be maintained, this seems an interesting case of diachronic emergence of new persons and properties.

11.5 Conclusion: But is this Really Emergence?

I hope in this paper to have shown that the opposition between mechanism and emergence is based on a misunderstanding, and that core features of emergent phenomena – dependence, autonomy, holism and novelty can be explicated in mechanistic terms. In addition, the mechanistic turn can shed light on differences in varieties of emergence. The distinction between mechanisms that produce vs mechanisms that underlie provides an analysis of the distinction between diachronic and synchronic emergence, and various interpretations of novelty, holism and autonomy can be shown to arise from different kinds of mechanistic organization.

I expect though that my proposed rapprochement between mechanism and emergence will be met with skepticism. One reason is that the mechanistic conception of emergence makes emergent properties the rule rather than the exception. It violates what Humphreys has called the rarity heuristic, which holds that any account “that makes ontological emergence commonplace has misidentified the criteria for emergence” (Humphreys, 2016, 54). But, as Humphreys effectively argues, there’s no clear argument for this heuristic, and we find ample evidence across the sciences for phenomena that can be characterized as both dependent upon and distinct from some base from which they emerge.

But I expect a deeper source of skepticism may lie in the way in which philosophers have understood the historical relationship between concepts of mechanism and emergence. Aristotle is often credited with being the first emergentist, because of his metaphysics of matter and form. Substances depend for their existence upon matter, but it is the way that this matter exemplifies form that gives a substance its novel properties and causal powers. But, the story goes, Aristotelian metaphysics was rejected in the scientific revolution and replaced by mechanical philosophy, an austere metaphysics in which all things in the world – at least all things in the material world – were nothing but matter and motion. Since that time, various

incarnations of this reductionist nothing-but-ism have thrived, from Laplacian determinism and de La Mettrie's "L'Homme Machine, to Wittgenstein's and Russell's logical atomism, to many modern articulations of microphysicalism in contemporary metaphysics. Humphreys calls the general strategy found in these sources **generative atomism** – because it assumes that there is some fundamental level of basic immutable entities, and that all higher levels of things are generated by dynamical and compositional principles or laws from these fundamental things. When Broad drew the distinction between mechanism and emergence by arguing that emergent phenomena cannot “be deduced from the most complete knowledge of the behavior of its components, taken separately or in other combinations, and of their proportions and arrangements in this whole” (1925, 59), it is clear that he was assuming an austere nothing-but form of mechanism.

Contemporary philosophical research on mechanism though has rejected this austere approach. The new mechanistic account is grounded in the practices of the life and social sciences. Those practices do not seek out a set of privileged atoms and properties from which all else is generated, but look instead at phenomena within a domain, and seek to identify particular kinds of entities, activities and interactions, and show how they are organized so as to be causally and constitutively responsible for their phenomena. As Bechtel and Richardson (2010, xliv) put it, some kinds of mechanisms exhibit emergent behaviors that are “neither weak nor epistemic.” Mechanistic phenomena are not simply generated by the activities of a set of immutable parts. They instead can (as discussed above) depend upon varieties of feedback and top-down causation (or something like it) whereby the whole influences the part.⁹ Investigation of these processes is, as Wimsatt has long argued, reductive, but not eliminative, and is piecemeal and local. The resultant ontology is, as he puts it, a rainforest ontology – with intricate dependencies, but also rich with novelty.¹⁰

The skeptic might still question how much distance mechanists can put between themselves and generative atomism. Given that novel entities and activities are going to be mechanism-dependent, just how novel can they really be? One way to pose this question is by appealing to the commonly held distinction between weak and strong emergence. As Jessica Wilson (2016, 2021) sees it, weak and strong emergence are both genuine ontological emergence, but represent different alternatives to traditional physicalism. Weak emergence follows the path of non-reductive physicalism, most commonly by appealing to the multiple realizability of

⁹Bechtel and Richardson's account is offered in the context of systems biology, but the conditions they require are generalizable to other domains. They argue that there are two necessary conditions for emergence. First, (some) interactions between components must be non-linear, and second, the state of the whole system must be characterized in terms of state-independent properties. These properties, which include properties of the whole system and its environment determine the properties of the components. While I haven't formulated conditions for emergence in just this way, Bechtel and Richardson's conditions seem to imply top-down causation and transformation, which are both inimical to generative atomism.

¹⁰These themes, which appear in much of Wimsatt's work since the 1970s, are nicely summarized in the introduction to (Wimsatt, 2007).

higher-level properties to justify claims of their autonomy. But on Wilson's view, weak emergence does not introduce genuinely novel causal powers. For that, one needs something like new fundamental laws to characterize the powers of the emergents.

There are, as I see it, three possible responses to such skepticism. First, one might grant that the kind of emergence compatible with mechanism is indeed weak emergence, but allow that there may be some rare phenomena – perhaps consciousness or agency – which exhibit strong emergence. A second approach is to argue that the strong emergentist position ultimately collapses into weak emergentism. The best and only kind of emergence we will find fails to introduce genuinely novel causal powers. The third and perhaps best response might be to question the basis for the distinction itself. The most common explications of strong emergence are cashed out in terms of levels, domains and laws. Novel causal powers will be expressed by causal laws characterizing relations between the novel entities' properties. But there are many strands of recent philosophy of science that suggest that analysis of laws and causes may not be the right way to go. Perhaps there are very few laws, or perhaps, as I prefer (Glennan, 2017), we should think of laws simply as descriptions of the behavior of mechanisms. If we think about laws and causes in this way, we may be forced to a different view of what would count as novelty, and mechanism might just provide novelty enough.

I have certainly not said enough here to convince skeptics that all emergent phenomena are within the reach of mechanism – and I am far from certain myself. But I hope I have said enough to show that the supposed incompatibility of mechanism and emergence reflects a misunderstanding of what mechanisms are. Mechanisms don't rid us of emergent phenomena; they show us how they work.

References

- Batterman, R. W., & Rice, C. C. (2014). Minimal model explanations. *Philosophy of Science*, 81, 349–376. <https://doi.org/10.1086/676677>
- Bechtel, W. (2017). Top-down causation in biology and neuroscience: Control hierarchies. In M. P. Paoletti & F. Orilia (Eds.), *Philosophical and scientific perspectives on downward causation* (pp. 203–224). Routledge. <https://doi.org/10.4324/9781315638577-12>
- Bechtel, W., & Richardson, R. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Reissue. MIT Press.
- Bishop, R., & Silberstein, M. (2019). Complexity and feedback. In S. Gibb, R. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence*. Routledge.
- Brigandt, I., Green, S., & O'Malley, M. A. (2018). Systems biology and mechanistic explanation. In S. S. Glennan & P. M. K. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy*. Routledge.
- Broad, C. D. (1925). *The mind and its place in nature. The mind and its place in nature*. Kegan Paul, Trench, Trubner & Ltd.. <https://doi.org/10.4324/9781315824147>
- Craver, C. F. (2007). *Explaining the brain*. Oxford University Press.
- Craver, C. F., Glennan, S. S., & Povich, M. (2021). Constitutive relevance & mutual manipulability revisited. *Synthese*, 199, 8807–8828. <https://doi.org/10.1007/s11229-021-03183-8>. Springer Netherlands.

- Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22, 547–563. <https://doi.org/10.1007/s10539-006-9028-8>
- Craver, C. F., & Kaplan, D. M. (2018). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, 0, 1–33. <https://doi.org/10.1093/bjps/axy015>
- Craver, C. F., & Tabery, J. G. (2016). Mechanisms in science. In E. N. Zalta (Ed.), *Stanford encyclopedia of philosophy*, Fall 2016.
- Dretske, F. (2021). A recipe for thought. In D. J. Chalmers (Ed.), *Philosophy of mind: Classical and contemporary readings* (2nd ed., pp. 351–358). Oxford University Press.
- Ellis, G. F. R. (2011). Top-down causation and emergence: Some comments on mechanisms. *Interface Focus*, 2, 126–140. <https://doi.org/10.1098/rsfs.2011.0062>
- Ellis, G. F. R., Noble, D., & O'Connor, T. (2012). Top-down causation: An integrating theme within and across the sciences? *Interface Focus*, 2, 1–3. <https://doi.org/10.1098/RFSF.2011.0110>. Royal Society.
- Fodor, J. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese*, 28, 97–115.
- Gibb, S., Hendry, R. F., & Lancaster, T. (2019). *The Routledge handbook of emergence*. Routledge. <https://doi.org/10.4324/9781315675213>
- Glennan, S. S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research*, 81, 362–381. <https://doi.org/10.1111/j.1933-1592.2010.00375.x>
- Glennan, S. S. (2017). *The new mechanical philosophy*. Oxford University Press.
- Glennan, S. S. (2021). Corporeal composition. *Synthese*, 198, 11439–11462. <https://doi.org/10.1007/s11229-020-02805-x>. Springer Netherlands.
- Glennan, S. S., & Illari, P. (2018a). Introduction: Mechanisms and mechanical philosophies. In S. S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 1–10). Routledge. <https://doi.org/10.4324/9781315731544>
- Glennan, S. S., & Illari, P. (2018b). Varieties of mechanisms. In *The Routledge handbook of mechanisms and mechanical philosophy*. <https://doi.org/10.4324/9781315731544>
- Glennan, S. S., Illari, P., & Weber, E. (2021). Six theses on mechanisms and mechanistic science. *Journal for General Philosophy of Science*. Springer Netherlands. <https://doi.org/10.1007/s10838-021-09587-x>
- Humphreys, P. (2016). *Emergence: A philosophical account*. Oxford University Press. <https://doi.org/10.1093/acprof>
- Kaplan, D. M., & Bechtel, W. (2011). Dynamical models: An alternative or complement to mechanistic explanations? *Topics in Cognitive Science*, 3, 438–444.
- Keller, E. F. (2008). Organisms, machines, and thunderstorms: A history of self-organization, part one. *Historical Studies in the Natural Sciences*, 38, 45–75. <https://doi.org/10.1525/HSNS.2008.38.1.45>
- Kim, J. (1993). The nonreductivist's trouble with mental causation. In *Supervenience and mind: Selected philosophical essays*. Cambridge University Press.
- Kistler, M. (2009). Mechanisms and downward causation. *Philosophical Psychology*, 6, 595–609.
- Krickel, B. (2018). *The mechanical world: The metaphysical commitments of the new mechanistic approach*. Springer.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Miłkowski, M., Clowes, R., Rucińska, Z., Przegalińska, A., Zawidzki, T., Krueger, J., Gies, A., et al. (2018). From wide cognition to mechanisms: A silent revolution. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.02393>
- Millstein, R. L. (2006). Natural selection as a population-level causal process. *The British Journal for the Philosophy of ...*, 57, 627–653.
- Newen, A., de Bruin, L., & Gallagher, S. (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

- O'Connor, T. (2021). Emergent properties. In E.N. Zalta (Ed.), *Stanford encyclopedia of philosophy*. Winter 2021.
- O'Malley, M. A., & Dupré, J. (2005). Fundamental issues in systems biology. *BioEssays*, 27, 1270–1276. <https://doi.org/10.1002/bies.20323>
- Paolini Paoletti, M., & Orilia, F. (2017). *Philosophical and scientific perspectives on downward causation*. Routledge.
- Paul, L. A. (2014). *Transformative experience*. Oxford University Press.
- Polger, T. W., & Shapiro, L. A. (2016). *Multiple realization book*. Oxford University Press.
- Putnam, H. (1975). The meaning of “Meaning.”. In *Minnesota studies in the philosophy of science philosophy of science. Vol. VII. Language, mind and knowledge* (pp. 131–193). University of Minnesota Press. <https://doi.org/10.1515/9783110871241.110>
- Sawyer, R. K. (2004). The mechanisms of emergence. *Philosophy of the Social Sciences*, 34, 260–282.
- Shapiro, L. A. (2000). Multiple realizations. *The Journal of Philosophy*, 97, 635–654.
- Silberstein, M., & Chemero, A. (2013). Constraints on localization and decomposition as explanatory strategies in the biological sciences. *Philosophy of Science*, 80, 958–970. <https://doi.org/10.1086/674533>
- Wilson, J. M. (2016). Metaphysical emergence: Weak and strong. In T. Bigaj & C. Wüthrich (Eds.), *Metaphysics in contemporary physics*. Brill.
- Wilson, J. M. (2021). *Metaphysical emergence. Metaphysical emergence*. Oxford University Press. <https://doi.org/10.1093/oso/9780198823742.001.0001>
- Wimsatt, W. C. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, 64, S372–S384.
- Wimsatt, W. C. (2000). Emergence as non-Aggregativity and the biases of reductionisms. *Foundations of Science*, 5, 269–297.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings*. Harvard University Press.
- Winning, J., & Bechtel, W. (2019). Being emergence vs. pattern emergence: Complexity, control and goal-directedness in biological systems. In S. Gibb, R. Hendry, & T. Lancaster (Eds.), *The Routledge handbook of emergence* (pp. 134–144). Routledge.
- Woodward, J. (2021). Downward causation defended. In *Top-down causation and emergence* (Vol. 439, pp. 217–251). Springer. https://doi.org/10.1007/978-3-030-71899-2_9

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Chapter 12

Emergence, Downward Causation, and Interlevel Integrative Explanations



Gil Santos

Abstract In this article, I propose a unified account of systemic emergence, downward causation, and interlevel integrative explanations. First, I argue for a relational-transformational notion of emergence and a structural-relational account of downward causation in terms of both its transformational and conditioning effects. In my view, downward causation can avoid the problems traditionally attributed to it, provided that we are able to reconceptualize the notion of ‘whole’ and that form of causality in a purely relational way. In this regard, I distinguish contextual or whole-to-part causation from downward causation, the latter defined by the existence of *second-order* structural relations. Finally, I argue that while emergence and downward-structural causation imply the in-principle failure of micro-determinism and therefore micro-reduction, they do not rule out the possibility of any type of explanation. On the contrary, they call for the development of interlevel integrative explanations.

Keywords Emergence · Downward causation · Integrative explanations · New mechanism · Relational ontology · Second-order relations

12.1 Introduction (to a Relational Ontological Approach)

According to the dynamic relational perspective that I will follow here, systemic emergence and downward causation must be conceptualized in terms of certain transformative and conditioning relations involving wholes, as *systems of relations*, and their proper parts, as *relata* of such relations.

A relational ontological view, as I conceive it, does not postulate that relations are all there is. Rather, it is an ontology according to which every particular entity (independently of whether it is conceptualized as an object, process, activity, event, etc.) owes its *identity* and *existence* to a relation between its *endogenous* and

G. Santos (✉)

Centro de Filosofia das Ciências, Departamento de História e Filosofia das Ciências,
Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal

© The Author(s) 2024

J. L. Cordovil et al. (eds.), *New Mechanism*, History, Philosophy and Theory of
the Life Sciences 35, https://doi.org/10.1007/978-3-031-46917-6_12

235

exogenous relations involving other entities, including in the context of higher-level relational systems (see Santos, 2015a: 439–442; 2020: 8693–8597).

In this sense, the basic ontological categories are not relations and objects but relations and *relata*. Objects are just one kind of *relatum*. Processes, events, and properties also relate with each other – causally, spatially, temporally, functionally, etc. Even relations can be themselves *relata*, for they relate and interact with each other in *structures*. As a matter of fact, many relations depend, in terms of their very *existence*, both on specific *relata* and *other relations*. For instance, gravitational interaction depends both on the existence of masses and relations of spatial distance. Finally, relations and *relata* should be seen as standing on the same ontological footing. Indeed, unless abstraction is involved, no *relatum* exists without being related to something, and no relation exists without being a relation between some *relata*. In this view, there is no room for absolutely intrinsic properties but only for endogenous and exogenous relational properties.

However, this correlation between relations and *relata* also includes the systems they form. In fact, *any relation between two or more relata immediately forms a system of which the relata are proper parts*. Therefore, relations, *relata*, and relational systems always come together as three co-relative ontological categories. From this vantage point, it makes no sense to ascribe an *absolute* ontological priority to *relata*, relations, or the systems they constitute.

This can also be seen from a temporal perspective. It certainly seems reasonable that there has never been a time in which different individuals existed without being related to each other in some way and without, therefore, being constitutive *relata* of some relational system. Likewise, there was never a time when systems existed without being themselves structured by the relations between their constitutive parts or *relata*. For this reason, it also makes no sense to assign an *absolute* temporal precedence to *relata*, relations, or their systems. From a temporal point of view, reality consists simply of the continuous generation of new relational systems from prior changes in other systems and relations and the ongoing transformation of entities both as *relata* and as relational systems.

This dynamic relational view thus rejects not only the metaphysical atomism or individualistic essentialism of old mechanistic philosophies, but also the holistic notion of brute or in-principle inexplicable emergent wholes. First, no whole exists apart from (and hence somehow independently of) the complete set of its parts' relations and its relations with the outside world. Second, the complete explanation of every system must always include an account of its formation as the historical outcome of prior transformations in other systems and relations.

I proceed as follows. In Sect. 12.2, I will present a *relational-transformational* account of systemic emergence, which was first elaborated in (Santos, 2015a) and later developed along the lines of a neo-mechanist perspective in (Santos, 2020).

In Sect. 12.3, I will articulate a *structural-relational* account of downward causation by answering the following four questions: what is a whole? (Sect. 12.3.1), what is the 'higher level' of an integrated whole? (Sect. 12.3.2), how should we conceptualize downward causation? (Sect. 12.3.3), and how does downward causation work? (Sect. 12.3.4). In this section, I will distinguish between two types of

causation: contextual or whole-to-part causation and downward causation. I define downward causation in terms of the existence of *second-order* structural relations, which is why I shall call it downward-structural causation.

In Sect. 12.4, I will show that it is the objective existence of systemic emergence and downward-structural causation that ultimately justifies the in-principle failure of micro-determinism and micro-reduction and that, at the same time, demands the use of *interlevel integrative explanations*.

Here lies, in my view, the *positive epistemological significance* of ontological emergence and downward causation. Furthermore, it is also here that we may find the real ontological and epistemological *novelty* of any neo-mechanistic perspective vis-à-vis the *old* mechanistic philosophies.

12.2 Emergence

The notion of systemic emergence has been historically defined in opposition to the notion of lower-level, whole-to-part or micro-reduction. Indeed, from the point of view of part-whole relations, there are only two possible ways of obtaining an absolutely *asymmetrical* or unidirectional relation of reduction: either *macro*-reduction or *micro*-reduction. That is, either we completely reduce the properties and respective relations (including laws) of the proper parts of a given system to the properties and laws *only instantiated* by that system, or we reduce the properties and laws of a system to the *intrinsic* or *system-independent* properties and respective relations (including laws) of its proper parts.

I italicized the phrases ‘only instantiated’ and ‘intrinsic or system-independent’ to highlight the fact that, in order to constitute a purely *asymmetrical* reduction, the reducing term must have *all* the resources required to account for the reduced term *independently* of the latter.

Consider micro-determinism and micro-reductionism. How completely micro-determined and therefore micro-reducible can a system’s property be when some of the parts’ properties contributing to its production are only instantiated *by virtue of* the parts’ integration within that very system? The only alternative form of reduction is a *partial* and *reciprocal* reduction. Yet, since the latter does not constitute an asymmetrical or unidirectional relation, it is compatible with some notions of emergence and downward causation.

My notion of relational-transformational systemic emergence (RTE) is defined in opposition to both complete micro- and macro-determinism and, consequently, to complete micro- and macro-reductive explanations. In my view, all processes of RTE are characterized by the necessary conjunction of two main features: (i) they all involve *relations*, and (ii) they all involve a *transformation* of the *actual* identity of lower-level entities as *parts* of some wholes or as constitutive *relata* of some systems of relations (Santos, 2015a, 2020).

By a change of the *actual* identity of an entity, I mean a change in the set of properties or behaviors that an entity actually has or manifests, even if by gaining

and losing some properties the entity will also change in terms of its *potential* identity, that is, in terms of the powers or capacities associated to such properties.

In this sense, I propose the following characterization of systemic emergence. A property *P* of a system *s* is emergent if, and only if,

- (i) *P* is a property of a specific global organization (*R*) of the proper parts of *s*, and
- (ii) *R*, and hence *P*, are *not* completely determined by (thereby *not* being fully explainable or reducible in terms of) the *intrinsic* or *system-independent* properties, and respective relations and laws, of the proper parts of *s*.

This means that the existence and explanation of property *P* depend on, at least, some *system-dependent relational* properties of the proper parts of system *s* – that is, properties which the lower-level entities only have or manifest *by virtue of* being *parts* of system *s*, or by virtue of being *relata* within the specifically organized system of relations called *s*. It should be noted that the qualifier ‘at least’ was included because *P* may also depend on some relations that the system *s* has with its external environment, including as a proper part of a yet *higher-level* system.

To say that *x* does not completely determine or produce *y* is the same as to say that *x* provides the *necessary* but *not the sufficient* conditions for the ontic determination or production of *y*’s existence or identity (see Bishop & Atmanspacher, 2006; Bishop et al., 2022: 27 and 94; and Santos, 2021:1).

RTE is found in that class of mereological complexes called ‘integrated systems’ (Bechtel & Richardson, 2010: 27), characteristically defined as being only ‘minimally decomposable’, i.e., whereby the ascription of independent behaviors or functions to their proper parts taken separately is impossible (Bechtel & Richardson, 2010: 26–31).

This notion of RTE is suggested by Wimsatt’s claim that emergence must “involve some kind of *organizational interdependence* of diverse parts” (Wimsatt, 1997: S375 – italics inserted; also: 2006: 673). In this sense, emergence implies not only a “dependence of a system property on the arrangements of the parts”, but also, and *above all*, “the context-sensitivity of relational parts’ properties to intra-systemic conditions” (Wimsatt, 2000: 270). In fact, it is important to distinguish between these two conditions and the different notions of emergence they imply. To say that an “emergent property is – roughly – a system property which is dependent upon the mode of organization of the system’s parts” (Wimsatt, 1997: S373), or that the “emergence of a system property relative to the properties of the parts of that system indicates its dependence on their mode of organization” (Wimsatt, 2006: 673), is not in itself sufficient to prove the inadequacy of a micro-reduction.

Emergent properties are not just organizational, collective, or non-distributive systemic properties. Indeed, metaphysical atomism has always recognized that many systems’ properties are dependent on specific modes of organization, combination, or arrangement of their parts. The issue is that for metaphysical atomists, any organizational property of a system is completely reducible to the *intrinsic* properties of its parts and their respective laws and causal relations. In this precise sense, organization is not enough. An organizational property can only be taken as a real emergent property if the *organization* is not itself completely determined by

the *intrinsic* properties and respective laws of the lower-level entities composing that organization (Santos, 2015a: 431–439; and Santos, 2020: 8690–8693).¹

The fundamental difference between RTE and mere ‘organizational emergence’ is that in the latter, intra-systemic relations and global organizations only intervene in the construction of the existence and identity of the *systems* taken as wholes, while in the case of RTE they also intervene in the construction of the identity, if not the existence itself, of the systems’ *parts*.

Some wholes may thus be said to be different from the mere sum of their parts, not only in the sense that they also depend on a specific organization of the parts, but, as Scott Gilbert put it, “in the sense that *the properties of each part* are dependent on the context of that part within the whole in which it operates” (2010: 618 – italics inserted).

In the important new introduction to the 2010 second edition of their book *Discovering Complexity*, Bechtel and Richardson clearly suggest the notion of RTE when identifying two basic conditions for a mechanistic notion of emergence that is “*neither weak nor epistemic*” (2010: xlv – italics inserted; see Santos, 2020: 8700–8701).

First, the activities or operations of the system’s parts depend on the actual behavior and the causal capacities of the other parts in a cyclic (non-sequential) and nonlinear way, and “to the extent that feedback is systemwide, these dependencies will result in operations that are specific to the system”. This condition refers to the fact that “the behavior of the components is system dependent” (2010: xlvi). Secondly, “the nonlinearities affecting component operations must in turn affect the behavior of the system” (2010: xlvi). As Bechtel and Richardson note, when these two conditions are met, “the systemic behavior is reasonably counted as emergent, even though it is fully explicable mechanistically” (2010: xlvi–xlvii).

In this regard, the idea that ontic emergence is “spooky” (Craver & Tabery, 2019) or “suspect” because it suggests a “discontinuity” between a system and its “parts, activities, and organizational features of the system in the relevant conditions” (Povich & Craver, 2018: 190) conflates features that should be distinguished so as not to render meaningless the very contrast between reductionist and interlevel integrative explanations. In fact, the ‘spooky’ or ‘suspect’ character of ontic emergence is just a consequence of ascribing an *absolute* meaning to the notion of ‘discontinuity’ between a system and its parts. Again, the question is whether parts considered in isolation, i.e., as *independent* individuals with all their alleged absolutely *intrinsic* properties, provide *both* the *necessary* and *sufficient* conditions for the

¹For an excellent depiction of atomism, both in its ontological and methodological aspects, see Humphreys (2016: 1–26). The notion of transformational emergence that I proposed in (Santos, 2015a) differs from the transformational perspective elaborated by Paul Humphreys (2016) in two essential ways. First, my notion of emergence applies to part-whole relations, thereby being explicitly based on a distinction between different levels of organization. Second, in my view, all transformation processes are caused by and must then be necessarily explained in terms of specific *relations*. This is the reason why I have been arguing for a *relational*-transformational notion of *systemic* emergence.

ontological determination of all properties and behaviors that they may exhibit in all possible relational contexts or systems, as well as for the ontological determination of all systems of which they can become parts.

Assuming the *realist* view that “the direction of explanation recapitulates the direction of determination” (Klee, 1984: 60), ontic systemic emergence simply means that we do not find both the necessary and sufficient conditions for the ontological determination and thus for a complete explanation of a given property of a system at the level of the *intrinsic* or *system-independent* properties of the parts of that system, and their respective relations.

RTE can occur, of course, either during the development of a system or in the generation of new systems. Furthermore, any system, just like any of its parts, also acquires properties by virtue of its exogenous relations with other systems, including as a *part* of or *relatum* within a further *higher-level* system of relations. As a matter of fact, it *is the existence of* such interlevel relations of determination that justifies the need for the explanatory task of ‘situating’ mechanisms or systems in their environments (Bechtel & Richardson, 2010). As Bechtel has stressed, the explanation of any mechanism always requires the consideration of “its relation to conditions in its environment” (Bechtel, 2011: 538), “including its incorporation within systems at yet higher levels of organization” (Bechtel, 2006: 40–41).

As it was said, RTE is defined in opposition to both complete micro- and macro-determinism and, consequently, to complete macro- and micro-reductive explanations. But to deny such forms of determination and explanation is *not* the same as to deny any form of determination and explanation. If the right relations are identified and the transformations they cause are taken into account, we can regard any emergent property as completely determined, and its complete explanation can be, at least in principle, provided. In Sect. 12.4, I will specifically address this issue when dealing with interlevel integrative explanations.

In the next section, I will show how RTE relates to downward causation, thereby clarifying the way in which parts can be determined by their wholes with the help of some concrete examples.

12.3 Downward Causation

Following the relational viewpoint presented here, I claim that there is only downward determination (causal or otherwise) if the *relations* which determine the parts of a given whole are at a genuinely *higher level* than those parts.

I construe downward causation (DC) as a particular form of downward determination because systems may determine their proper parts by means of both *causal* and *non-causal* relations. For example, parts may acquire or lose some causal powers without undergoing any causally induced inner structural changes if such powers are to be considered genuine extrinsic relational properties; parts’ behaviors may be constrained by the topology of their systems’ structures, and so on.

The parts of a system may alter their *actual* identities or only their *potential* identities as a result of downward determination. By a change of the actual identity of an entity, I mean a change in the set of properties or behaviors that it actually has or manifests, even if, by gaining and losing some properties, the entity will also change in terms of the powers or capacities associated with such properties. By a change of the potential identity of an entity, I mean a change in its set of powers or capacities (without considering their actual manifestation) or in terms of a reduction or extension of its degrees of freedom.

The downward causal determination of the actual identity of the proper parts of a system may be called *downward causal transformation*. The downward causal determination of their potential identity may be called *downward causal conditioning*. Furthermore, since parts may be changed by acquiring or losing powers, they may be subject to downward causal conditioning, either in *empowering* or *constraining* terms (Archer, 1995; Hooker, 2013: 761).

Finally, because DC is a causal relation, it must necessarily be seen as a *diachronic* process. This means that even downward causal conditioning is never, strictly speaking, synchronic. To say that part x of system s is conditioned by s at time t is to say that x can only act this way or that way at any time *after* t , thereby changing the set of its possible future behaviors. For example, to say that part x has lost a given power or capacity P at time t is to say that x cannot act in a P -way *from that moment on*, i.e., after t . At any time t , every entity is *just acting* given the possibilities defined *before* t . The effects of causal conditionings thus always come *after* the imposition of such conditions.

The notion of DC has two well-known problems. The first problem is that it contradicts the principle of causal closure or completeness of the micro-physical level of reality and its associated principle of overdetermination. The second problem is that it seems to contradict the notion that causal relations must be non-reflexive. I will address these issues by answering the following four questions: what is a whole? (Sect. 12.3.1), what is the ‘higher level’ of an integrated whole? (Sect. 12.3.2), how should we conceptualize DC? (Sect. 12.3.3.) and how does DC work? (Sect. 12.3.4). In the following, I propose a structural-relational account of DC, which can avoid the problems traditionally attributed to it as well as allow it to be easily placed within a neo-mechanistic framework.

12.3.1 What Is a Whole?

The main reason for the troubles that the notion of DC faces when confronted with the classical notion of causal relations as *non-reflexive* lies, in my view, in the very notion of ‘whole’ that has been (more or less tacitly) adopted.

If a whole, in its broader sense, is just a *set* of parts and their respective relations, how can that *set* causally affect its *subset* of parts? According to the received view, causal relata must be spatially and temporally distinct, and thus not related compositionally. Part-whole relations, in turn, should be limited to compositional

relationships, meaning those of ‘constituting’ and ‘being constituted by’, either in purely spatial or mechanistic terms (working parts, component operations, etc.). Thus, how can a whole causally influence or condition its constituent parts?

In my view, part-whole causal relations should not be seen as relations in which a whole, taken as a *set* of parts and their respective relations, causally interacts with, thereby affecting, the *subset* of its parts. Their causal relations should neither be thought of as relations between two object-like things: the whole as an object *vs.* its parts as a plurality of objects. Both views are conceptually flawed, thereby creating unnecessary problems.

If by the term ‘whole’ (or system) we simply mean a set of elements as proper parts and their respective relations, then any whole has two correlative but distinct dimensions: the set of its proper parts and the set of its proper relations. Therefore, the proper parts of a whole are just the *relata* of the *relations* constituting that whole. In this view, *any relation between two or more relata immediately forms a whole of which the relata are proper parts*. In short, a whole is to its *proper parts* what a system or network of relations is to its constitutive *relata*.

Of course, there are different kinds of wholes, depending on the nature of their constitutive relations, the degree (if any) of their organization and interdependence, their stability or persistence conditions (some are highly transient, while others are very stable), etc. But the above characterization stands as the most general notion of a whole in relational terms.

How can then wholes causally interact with their parts? To address this question, I find it helpful to distinguish between two different *perspectives* of a whole. I shall call them the *outside* and the *inside* view. From the *outside* view of a whole, a whole is just a set of entities, as its proper parts, and their relations. From the *inside* view of a whole, we should say that for every part taken as a *relatum*, *its* whole is just the set of relations among *all its co-relata*. In this sense, a whole is nothing more than the ‘relational context’ in which an individual is embedded.

To ask, then, whether a part is affected by its whole is just to ask whether an individual entity is affected *by being a relatum* in a given *system* or *network of relations involving all its co-relata* (Santos, 2015b). This is to say that for each part, there is a different whole taken as a system of relations among all other co-relata. Only the *outside* view ‘presents’ a unique whole.

We can thus make sense of whole-to-part causation without invoking the holistic notion of a whole as an unanalyzable individual or primitive thing. To be part of a whole is simply to be a *relatum* within a specific system of relations. Thus conceptualized, whole-to-part causation is no longer a reflexive relation.

As a matter of fact, this relational perspective has already been advocated by Jean Piaget, in 1950, while addressing the relationships between sociological and psychological explanations (a problem classically polarized by the holistic and individualistic views). In the context of that analysis, Piaget poses the following question: “If the individual is the element and the society is the whole, how is it possible to conceive a totality which modifies the elements which make it up, without making use of other material than these elements themselves?” (1995: 39). Piaget’s answer was that the notion of social totality, or whole, should not be conceived as

“a combination of pre-existing elements”, nor as a “novel entity” (in the sense of something existing over and above its parts), but as “*a system of relationships*, each of which in its own right brings about a *transformation* of the elements thus related” (1995: 41 – italics inserted).

It is only because it is often assumed (even without full awareness) a holistic, mystifying notion of whole, that whole-to-part causation is frequently seen as a highly unique type of relation endowed with a certain air of mystery. As soon as we demystify the notion of whole, we can easily see how widespread whole-to-part or contextual causation is in nature.

12.3.2 What Is the ‘Higher Level’ of an Integrated Whole?

Let us then assume that it is not the whole as an individual thing or *object* that can causally affect its parts taken as a collection of further objects. Instead, the *system of relations* constituting a whole causally affects each of its constitutive *relata* as proper parts of that system. But in a system of relations, what stands at a *higher* level than the *relata* constituting that system? In my view, the higher level of any non-aggregative system is simply the level of the *global organization* of all its proper parts, as well as the properties and laws of that organization taken in and of itself.

A distinction must nevertheless be made between ‘component systems’ and ‘integrated systems’ (Bechtel & Richardson, 2010: 26–27). The organizing relational structure of a component system is constituted by merely quantitative and combinatorial relations between quasi-independent parts. On the contrary, the organizing relational structure of any integrated system is constituted by some qualitatively *transformative* and *interdependent* relations between its proper parts (see Santos, 2020: 8697–8700).

The fact that the relevant inter-individual relations within a given system are *interdependent* means that they do not occur or develop independently of each other, which means that they are not conceivable as separate or atom-like dyadic relationships. Relations do not come one by one, acting separately from the others, and affecting one-after-another each *relatum* at a time. Integrated systems are systems of interdependent parts, which means that their behaviors and relations are also interdependent. This is why integrated systems’ parts are said to be only ‘minimally decomposable’, as it is impossible to ascribe ‘independent’ properties or causal works to them (Bechtel & Richardson, 2010: 27, 31).

But it is not only that the inter-individual relations between the parts of an integrated system are dependent on each other. The key feature is that their interdependence follows specific *system-level modes of organization*. That is, parts’ relations are not dependent on each other in a purely haphazard, contingent, or arbitrary way. For example, in eukaryotic cells, protein folding always takes place *after* translation; transcription always takes place *before* translation; mature mRNAs are always translated *outside* of the nucleus where this process always involves the causal

intervention of ribosomes in *another region* of the cell called cytoplasm. All these inter-individual relations are not dependent on each other solely in terms of their *doings* and *outcomes*; they also follow a *system-specific global order or organization*. In sum, the higher level of any integrated system is made up of two different kinds of *second-order relations*, namely, relations of *systemic interdependence* and relations of *lawful interdependence*.

Now, these specifically organized systems of relations are at a clearly *higher level* of organization than the *inner* organization of *each* of their constitutive relata as lower-level subsystems. Furthermore, the mode of organization of any integrated system – which always involve specific re-equilibration or self-regulation causal processes (homeodynamic and allodynamic), as well as topological and temporal relations – is always *new* and *different* from the inner modes of organization of its parts.

12.3.3 *How Should We Conceptualize Downward Causation?*

From what has been argued, it follows that DC must be conceptualized as referring to the set of transformations and conditionings that a *specifically organized system of interdependent relations* exerts in each of their relata as lower-level parts of that system.

Therefore, DC implies the existence of *second-order relations* that structure or organize in a *specific way* (causally, spatially, and temporally) the *first-order* relations between the parts of a system, thereby defining the *way* in which these first-order relations determine and change the parts. In other words, DC does not apply to cases where an individual is just causally affected by a set of first-order relations with other individuals but to cases where individuals are causally affected by the *way* their first-order, inter-individual relations are *themselves related* (structurally and functionally) in a *systematic* manner.

This only reinforces the need to distinguish the *macro-relational structure* of a system from its *micro-compositional structure*, i.e., from the set of all properties and first-order, inter-individual relations among the system's parts (Santos, 2020: 8697–8698).²

Some first-order relations can only exist if organized in a particular way (can one imagine translation occurring before, rather than after, transcription in any cell?).

²As Peter Simons has written, “[a structure is] the *total* or *overall* relation of all the parts, as distinct from the multitude of binary and other relations between and among the parts. In fact, we need even to consider the various kinds and properties of the individual parts as part of this overall relation or pattern. In algebra or model theory, it is usual to define a ‘relational structure’ as a *sequence* $\langle D, R_1, R_2, \dots \rangle$ where D is the domain (set of parts) and R_1, R_2 , etc. the specific relations among these elements. The structure is here as it were a sort of *super-relation* on the domain and relations” (1987: 355–356). For an in-depth and comprehensive understanding of the notion of structure and its key role in the formation of contemporary science, see Piaget, 1971b.

In some systems, certain relations can be organized in different ways (within certain limits, of course), but their causal effects will also be different. In any case, what *causally affects* (transforms and conditions) each part of an integrated system is not a sum of independent inter-individual relations but a *specifically organized set* of them. To put it differently, in integrated systems, the *causal workings* of inter-individual relations cannot be separated (except through abstraction) from *how* they are specifically structured or organized. The specific modes of organization (i.e., relational structures) of integrated systems must thus be counted as real contributing *causes* of the behavior of those systems' parts. Inter-individual causal relations and their distinctive modes of organization via some second or higher-order relations do not operate separately; they *come together, work together and act together*.

It should go without saying that the concept of causation associated with second-order structural relations cannot be understood in terms of the concept of *efficient* causation for the very obvious reason that the latter was designed to deal exclusively with first-order, inter-individual relations. While all inter-individual causal relations act as *efficient* causes, their specific modes of organization and interdependence act as *downward structural* causes (see Lawson, 2013: 287; and 2019: 38, 87–88, 199–200, 214–219).

A further distinction may be worth emphasizing. While in contextual or whole-to-part causation, the whole that causally acts on each of its parts can be said to constitute a mere *plurality* (*viz.*, the set of *all other* parts' relations; see above, Sect. 12.3.1), in the case of downward-structural causation, the whole constitutes a genuinely new *individual* due to the structural and functional *unity* (or interdependence) between its constitutive relations. As Simons has noted, it is important “to distinguish between a collection of many individuals and the one individual they compose, if they do” (2006: 599, n. 4). However, as Tony Lawson rightly pointed out, it is the ‘organizing relational structure’ of a composite whole, rather than the whole-as-a-whole, that may be said to exert top-down or downward causation (2013: 287; 2016: 431–432; 2019: 38, 74 n. 9, 214–219).

From this perspective, it is possible to discern the occurrence of DC in any integrated system, whether hierarchically or heterarchically organized. Indeed, even in the most strongly hierarchical systems, the ultimate ‘master controller’ (so to speak) of the parts' behaviors is always the specifically organized set of relations that structure those systems. In the last instance, it is never a specific part that downwardly causes the other parts. The real agents of DC are always specifically organized systems of relations because it is ultimately *by virtue of* them that some part may have a more relevant causal role in determining or regulating the activities of the other parts.³

The same goes for systems where power or control is more equitably distributed. Even though the powers that each part has are obviously powers *of* each of these

³For a distinction between different notions and uses of the concept of ‘hierarchy’ in systems and network neuroscience, see Hilgetag and Goulas (2020). As the authors show, “diverse ‘hierarchical’ concepts lead to different interpretations of the empirical data, with diverging functional implications” (2020: 8).

parts, the ultimate *source* of their instantiation is always a function of the interplay between each part's inner structure and the organized set of relations that this part has with all the other parts as its co-relata. Indeed, only in a fictional world of abstract individuals would entities have or lack powers exclusively in terms of their inner structures.

12.3.4 *How Does Downward Causation Work?*

I argued that DC should be conceptualized as the set of transformations and conditionings that a specifically organized system of interdependent causal relations exerts on each of their relata as lower-level parts of that system. Furthermore, as I noted above, because DC is a type of causal relation, it must necessarily be viewed as a *diachronic* process. At any given point in time, the organizing relational structure of a system affects or partially determines the behaviors that the parts of that system will or may instantiate at a later time. Likewise, at any time t , every entity is just acting given the *possibilities* defined *before* t .

The same is true of upward causation. The individual behaviors of each system's part *causally* contribute to the maintenance or modification of the collective behaviors of the system (Santos, 2015b), thereby contributing to the 'reproduction' or 'transformation' of the system's structure (Archer, 1995; Lawson, 2019). But at any given point in time, the behaviors of the parts simply "collectively constitute (along with the relevant organising structures)" the behaviors of their system (Lawson, 2019: 217).

Therefore, parts do not change because the wholes they compose change, as a mere mereological consequence of being parts of such wholes (Craver & Bechtel, 2007). Parts change because they are relata within specifically organized systems of interdependent transformative and conditioning relations.

This is well exemplified in self-organization processes. Properly speaking, self-organization is a process by which a system *reorganizes* itself in terms of the relations between its parts as a result of some external disturbances that threatened the original organization. If the reorganization process succeeds in 'assimilating' such disturbances, the system will then persist (Atlan, 1979: 165–170; 2011a, b). Rayleigh-Bénard convection, but also immune systems, are two of the most well-known and studied examples of such dynamics (Atlan & Cohen, 2006; Atlan, 2011a, b; Bishop, 2008; Bishop et al., 2022: 37–43; Cohen et al., 2016). The new systemic properties generated by self-organizing processes are thus emergent in a relational-transformational sense (see above, Sect. 12.2).

Most of the behaviors that entities exhibit as lower-level parts of integrated systems can only be explained by the fact that they are parts of such systems, thereby being determined by the structural organization of their constitutive relations. In any organized system of interdependent causal relations, the effects are propagated or transmitted in a specifically ordered manner, with each part thus being both *directly* and *indirectly* related to all other parts' relations in a system-wide way.

For example, although it takes DNA and RNA to produce proteins, it takes proteins (e.g., transcription factors) to regulate the activity of DNA and RNA. Proteins, in turn, will play a key role in manufacturing (e.g., RNA polymerases) the very nucleotide sequences that code for specific sequences of amino acids from which new proteins will then be produced. When two integrated system's parts interact, they are of course the direct causal agents of their own interaction. Yet that relation is *directly* and *indirectly* related to the globally organized set of relations involving the other parts of that system. For example, in eukaryotic cells, the interactive process of translation directly involving mRNAs, tRNAs, and ribosomes is *indirectly* but necessarily related to the outcome of the transcription relation, which directly involves DNAs and RNA polymerases, as well as all other subsystems which contribute to the editing and correction of transcription errors (Vecchi, 2020a).

Consider the well-known causal contribution of chaperons to the protein folding process. Polypeptides and chaperons are, of course, the direct causal agents or interactants of their own relation, but that relation itself depends, both *directly* and *indirectly*, on many other cell-specific types of relations. It is, of course, understandable that when highlighting the importance of the specific causal contribution of chaperons to the process of protein folding, we limit our analysis to *their* causal interaction. When focusing on, and thus conceptually abstracting, that *pair* of interactants, it seems that it is *all about them*. But that can only be done at the cost of a necessary, but highly selective, abstraction, leaving outside many other intra-cellular causal factors (e.g., water molecules, prosthetic groups, osmolytes), without which the relation between chaperons and polypeptides would *not take place* (Santos et al., 2020) – as a matter of fact, without which those proteins would *not even have come into existence*.

This kind of dynamic illustrates the *highest* degree of individuals' dependence on specifically organized systems of relations, that is, a dependence not only in terms of their behaviors or identity but in terms of their own *existence*. No individual comes into existence without having been generated by some system of relations. No individual can persist independently of the interplay between its endogenous and exogenous relations with a specific environment. And many individuals cannot even exist and persist except as parts of specific systems. This is what happens in the particular subclass of integrated systems that Richard Levins has called 'evolved systems', that is, systems "in which the component subsystems have evolved together" (1970: 76).

For example, ribosomes and mitochondria cannot persist as functional structures outside of cells. DNA molecules can but at the cost of becoming just one of "the most nonreactive, chemically inert molecules in the living world" (Lewontin & Levins, 2007: 239). To say, then, that some polypeptides only acquire their native structure by virtue of their relations with other proteins, such as chaperons, means that some *products* of the cell system (i.e., polypeptide chains) only acquire their native structure by virtue of some causal interactions with other *products* of the cell system, called chaperons. Both *relata* *are* what they are and *act* and *interact* the way they do by virtue of being *constructs* and *relata* of specifically organized systems of relations called eukaryotic cells, involving other *relata*, such as proteins, DNA, and RNA molecules.

This means that the *inner* organizational structures of the lower-level entities provide only the necessary but not the sufficient conditions for the ontological determination of the properties, behaviors, and causal powers that they have and actually manifest in the context of different systems of relations. And this also means that, even though inter-individual relations may be the only *empirically observable* relations, we cannot stop the explanation at the level of such relations when we know that these do not come into existence and take place as separate things but are systemically interdependent in a specifically organized manner.

Unsurprisingly, this is a much-debated topic in contemporary theories of sociological explanation. The idea of stopping the explanation of an integrated system, or part of it, at the level of its empirically observable inter-individual relations corresponds to a “flat” ontological view, where “networks remain linkages between nodes instead of networks of relations” (Donati & Archer, 2015: 22), “despite there being no such thing as context-less action” (2015: i). In the light of this new *atomism of relations*, social explanations should only involve “interpersonal relations (the Individualist concept of ‘social structure’)”, as “the social context should be reduced to refer to nothing but ‘other people’” (Archer, 1995: 36, 34).

As Auyang observes, in “advocating the reductive elimination of social concepts, [methodological individualism] mistakes situated individuals for bare individuals and overlooks the causal feedback that society has on individuals”. In particular, it “forgets that citizens are not Hobbesian men mushrooming from the earth; even in their most self-centered mode, they have *internalized* much social relation and conditioning, so that social concepts *have been built into* their characterization” (Auyang, 1999: 121 – italics inserted).

This is also the reason why most if not all properties, actual behaviors, or interactions of integrated systems’ parts cannot be explained in terms of absolutely *intrinsic* potentialities or dispositions.

A typical example is DNA’s property of ‘being a unit of inheritance’ or the property of ‘being a gene’, when conceived as the causal power of a genetic sequence ‘to code for a particular chain of amino acids’ and ‘to contribute to the construction of functional phenotypic traits’ (Santos, 2020: 8703–8705). These causal powers are system-dependent relational properties that DNA molecules and nucleotide sequences acquire only by virtue of interacting with some other relata, such as RNAs and proteins – including the existence of quality control mechanisms that successfully edit transcription errors – in the context of a *specifically organized* set of transformative and conditioning first-order relations, such as transcription, splicing, translation, and protein folding (Strohman, 1997; Shapiro, 2009; Vecchi, 2020a). This is a clear example of the *empowering* effects that downward causal conditioning can have (see above, Sect. 12.3).

Furthermore, the very definition of what a gene is “depends on the properties of the cell in which the DNA is embedded”, since the properties of a cell “are at least partly determined by transcription of DNA, but, in turn, cellular properties also determine which sequences are to be transcribed, in which combinations, and in what order” (Keller, 2010: 30; see also Atlan & Koppel, 1990). As Keller has stressed, “the necessary dependency of genes on their cellular context, not simply as

nutrient but as *embodying causal agency*, is all too easily forgotten” (Keller, 2001: 309 – italics inserted). This leads to the notion that the findings of developmental biology “point neither to cytoplasmic nor to nuclear determination but rather to a complex but highly coordinated system of regulatory dynamics that operate simultaneously at all levels: at the level of transcription activation, of translation, of protein activation, and of intercellular communication – in the nucleus, in the cytoplasm, indeed in the organism as a whole” (Keller, 1995: 29–30).

However, it is still not enough to consider the overall inner structure of an organism, for “the present environment and its history, at the scales of the cell, the person, the group and the biosphere, interact with the genome to determine its expressions and effects” (Cohen et al., 2016: 6). Moreover, “the contribution of a gene to a phenotype cannot always be separated from the contribution of the environment, despite sophisticated calculations, because the interactions between genome and environment are not linear and not additive” (*Idem*). The whole formation of the structural and functional identity of any eukaryotic cell constitutes a prototypical example of a process of downward-structural causation with the intervention of multiple system-wide feedback loops involving both intra- and extra-cellular interactions (Santos, 2020: 8703–8705).

Another example is provided by the fact that the macrostructure and function of proteins in their native or post-folded structure cannot be accounted for solely in terms of the system-independent properties or potentialities of the components of the primary structure as if they were essentially immutable entities (Santos et al., 2020). Polypeptides only acquire some of their potentialities and functions by undergoing a series of *structural transformations* (i.e., the acquisition of their so-called primary, secondary, tertiary, and quaternary structures) through the developmental process of folding by interacting with specific environmental inputs (e.g., pH and temperature) and contingently present substrates (e.g., water molecules and prosthetic groups) in specific cellular and organismal systems. Higher-level cellular, organismal, and environmental systems of relations do not thus merely trigger the manifestation of some potentialities already given *ab initio* but actually play a causal role in the very generation of new powers or capacities (2020: 377–380). This is a clear case of downward causal transformation and conditioning. And this is the reason why there are good reasons to support a relational-construction-based view of protein development and potentialities formation, which in turn requires the analysis of the dynamical interplay between lower- and higher-level organized systems of relations (2020: 363).

The notion of ‘developmental potential’ might be used as another illustrative example. Even though organisms are the units of development, the potential for that development does neither lie entirely in themselves nor in a specific part of them (such as their genomes). The extra-organismal environment must be counted as one of the three necessary, partial, and complementary causal bases for development potential (Vecchi & Santos, 2023). Therefore, if the genome, the developing organism, and the extra-organismal environmental materials are to be counted as proper structures of the causal basis for an organism’s developmental potential, the latter is not a given. Rather, it is the result of an interaction-based construction, a process

sometimes generating genuine developmental novelties. Hence, what would seem, and is indeed often assumed to be, an intrinsic potential or disposition, is in fact a multi-causal-based extrinsic relational potential of organisms *constructed* in the course of their own development (2023: 26).

This is the reason why we ought to endorse a dynamic-constructivist view of developmental potential, as phenotypes are often constructed out of biotic and abiotic environmental materials. As West-Eberhard notes, “due to changes in both genomic and environmental inputs” (2003: 13), as well as because many of the structural and functional changes undergone by the developing organisms are caused by the assimilation, functional integration, and deployment of environmental resources (a process which West-Eberhard calls ‘developmental entrenchment’), “developmental potentialities” themselves “change” during ontogeny (2003: 13 and 500 ff.).

In the following section, I will elucidate how systemic emergence and part-whole relations of reciprocal and partial co-determination necessitate the use of interlevel integrative explanations in a way consistent with a neo-mechanistic approach.

12.4 Interlevel Integrative Explanations

12.4.1 *The Birth of a ‘New Mechanism’ and Its Integrative Explanation Models*

The birth of a neo-mechanistic view in the twentieth century was essentially due to the impact that cybernetics had – particularly on the biological sciences – from the 1940s onward. The real *novelty* of this neo-mechanistic view relative to the *old* mechanistic philosophies can be primarily found in the recognition of a new, systemic form of causality (typically involving cyclic, feedback, feed-forward, and non-additive relations), and in the subsequent overcoming of the most simplistic notions of reduction in scientific explanation.

In this sense, it is easy to understand why the advent of a new mechanistic approach was seen as very good news for all those looking for a naturalistic way to overcome both the *neo-vitalist* and *old* mechanistic views in biological theory.

Norbert Wiener was explicit in recognizing the birth of a new mechanistic view in his 1948 book *Cybernetics*. According to Wiener, the creation of the modern automata represented both the “complete defeat” of Vitalism (indeed, “the whole mechanist-vitalist controversy has been relegated to the limbo of badly posed questions”) and the birth of a new, non-Newtonian mechanistic view in biology. Yet, this “new mechanics is fully as mechanistic as the old”, since “the essential mode of functioning of the living organism” is basically “the same” as that of the modern automaton (Wiener, 2019: 62–63, also: 54).

As Piaget observed in his 1967 *Biology and Knowledge*,

just at the time when biology was freeing itself from its restricting mechanistic ideas, and when some thinkers, confronted with this deficiency in traditional physical causality, were toying with the idea of a return to vitalism and finality, *a complete reelaboration of the mechanistic approach* opened up new perspectives along lines which corresponded exactly to those notions of circular or feedback systems or of cyclic rather than linear causality (Piaget, 1971a: 130–131 – italics inserted).

Regarding this neo-mechanistic approach (new in relation to “the mechanistic approach of old-fashioned physics”), Piaget highlights the importance of Cannon’s notion of homeostasis and, in general, the “rethinking of causality along the lines since followed by cybernetics”, which, in turn, allowed the scientific study of “auto-regulatory” systems, and “an extension of the general idea of organization, seen as a system of transformations” (1971a: 129–131).

In his 1972s paper, “Noise as a principle of self-organization”, Henri Atlan also acknowledged that the onset of cybernetics in the late 1940s prompted the birth of a “new mechanism” that “progressively imposed itself on biology” (2011b: 95–96). Atlan emphasized the discovery of many neo-mechanistic properties, such as the “redundancy of components, redundancy of functions, complexity of components, delocalization of functions”, “adaptability” and “self-organization” (2011b: 96–98).

Now, with this *broadening* of the scientific concept of causality came a *rehabilitation* of causal models of explanation as well, which represented an overcoming of the empiricist prejudices of both positivism and neo-positivism (see Bunge, 1959). Some of the neo-mechanistic views that would be developed throughout the 1950s, 1960s and 1970s are, of course, in some crucial respects different from the views which would be promoted, from the 1990s onwards, by the so-called “Chicago Mechanists” (Wimsatt, 2018). For example, according to both Piaget and Bunge, although causal explanations are a necessary step of scientific research in the non-formal sciences, they necessarily *depend* on the *previous* discovery and coordination of laws as statements of general facts or repeatable relations. Accordingly, causal *explanations* – unlike the *search* for the causal relations based on which one may then explain – are necessarily *deductive*, as they always proceed (as Aristotle has put it) from the more general to the more particular (e.g., Piaget, 1950a: 265–341; 1963; 1967a: 766–772; 1970a: 47–49; 1970b: 233–234; 1971b: 37–44; Bunge, 1964, 1967: 3–65; , 1983: 3–16). Nevertheless, in spite of such differences, their perspectives on inter-level explanations largely coincide.⁴

⁴For an overview of this ‘revival’ of causality and causal explanations since the 1950s, with a special emphasis on the development of physics, biology, psychology, and the social sciences, see Bunge (1982). From the mid-1970s and 1980s, numerous other authors contributed to the rehabilitation of causal models of explanation, including Rom Harré, Roy Bhaskar, Michael Scriven, Peter Railton, Paul Humphreys, and Wesley Salmon. For a selective overview of this movement, though limited to the philosophical literature, see Salmon (1989).

Drawing on his work on developmental cognitive psychology and developmental epistemology, Piaget was one of the first scientists to explicitly recognize the need for an interlevel integrative model of explanation as an alternative to the reductionist models of explanation, equally supported by positivist and classical mechanistic philosophies. Piaget named his alternative model of explanation, ‘*reciprocal assimilation*’, ‘*reduction by interdependence*’, or ‘*hybridization*’ (1950b: 64–79; 1967b: 1151–1182 and 1249; 1970a: 46; 1970c: 469, 525).

According to Piaget, three types of dependence relations among theories addressing different levels of organization can be defended:

- (i) reduction from the ‘higher’ to the ‘lower’;
- (ii) irreducibility of the phenomenon of the ‘higher’ level; and
- (iii) reciprocal assimilation by partial reduction of the ‘higher’, but also by enrichment of the lower by the higher (1970c: 469).

In the latter case, “a more complex science can be integrated into a simpler one, but then it *enriches* the latter to *transform* it into a new system through the *interdependence* of the superior and the inferior” (Piaget, 1967b: 1182).

For Piaget, “even in physics attempts to reduce the complex to the simple, for example, electromagnetic to mechanical phenomena, lead to syntheses in which the more basic theory becomes *enriched* by the derived theory, and the resulting *reciprocal assimilation* reveals the existence of structures as distinct from additive complexes” (Piaget, 1971b: 45).⁵ In this line of thought, Piaget would go as far as writing that one can “be quite relaxed about the prospect that living phenomena will one day become reduced to physico-chemical ones; here, as in physics, reduction will not mean impoverishment but such *transformation* of the two terms connected as *benefits both*” (Piaget, 1971b: 45–46 – italics inserted). In other words, “if a physico-chemical explanation of life can be expected, our present physico-chemistry will gain *new properties* thereby, thus becoming more ‘general’ instead of being applied exclusively to more and more special fields” (1970c: 469 – italics inserted; also: Piaget, 1950b: 75–79).

A similar perspective was defended by Monique Lévy (1979) in the context of her analysis of the relationships between biology, chemistry, physical chemistry, and physics. Unlike reduction, taken as a purely “asymmetric relation”, *reduction by synthesis* “assigns specific theoretical roles to each discipline” (Lévy, 1980: 152–153). According to Lévy, as “a consequence of the specific role played by each of the disciplines partaking in the reduction, every ‘reduction by synthesis’ proceeds not by annexation of a domain in the frame of another, but by interaction, and by reciprocal enrichment” (Lévy, 1980: 153). For example, addressing the alleged reduction of chemistry to quantum mechanics, Lévy noted: “If we limit chemistry

⁵ Piaget presents as paradigmatic examples of these processes of reciprocal assimilation the simultaneous ‘geometrization’ of gravitation and ‘physicalization’ of the Riemannian space curves in Einstein’s general relativity theory, as well as the interactions between mechanics and electromagnetism, which, after a period of attempts at unilateral reduction, would then lead to the creation of wave mechanics (1967a: 768; 1970a: 46).

to what could be deduced from physics alone, whole areas of this science would disappear (kinetics, organic chemistry, biochemistry, non-equilibrium processes)” (1979: 348).^{6,7}

Piaget’s model of reciprocal assimilation (from 1950 onward), Lévy’s notion of reduction by synthesis (1979/1980), or Bunge’s views on integration (e.g., , 1983: 31–45 and 165–175) are precursor variants of what is presented today as ‘interlevel integrative explanations’ (e.g., Bechtel, 1986a; Craver, 2005; Brigandt, 2010; Craver & Darden, 2013: 161–185), ‘multiscalar’ or ‘multi-level contextual explanations’ (Bishop et al., 2022), or simply ‘interdependence’ and ‘hybridity’ (Cat, 2022).

In Craver and Darden’s (2013) classification of the different ways an integrative explanation can occur in a mechanistic explanation, ‘interlevel integration’ fulfils a special role. It consists in the integration of what different fields find at different levels of organization, “either by looking up to see how a phenomenon is integrated within higher-level mechanisms or by looking down to see how a phenomenon is integrated with lower-level mechanisms” (2013: 163). As Craver and Darden note, “many of the great achievements in the history of biology involve bridging different levels of mechanisms” (Craver & Darden, 2013: 167). Interlevel integration thus represents an alternative form of explanation to the classical micro- or lower-level reductionist models, and to their associated idea that fields studying lower-level phenomena are “always more fundamental in explanations” (Brigandt, 2010: 297).

Yet there is more to integration than simply putting forward integrative theories. Aside from integrating explanations, “integrating methods (inference and modeling methods as well as experimental methods) and integrating data” are required (Brigandt, 2013: 463). Philosophy thus needs to understand “how concepts, methods, and explanatory resources are in fact coordinated, such as in interdisciplinary research where the aim is to integrate different strands into an articulated whole” (Love & Lugar, 2013: 548). For example, evolutionary developmental biology, which is an attempt to promote a theoretical integration of evolutionary biology and developmental biology, “faces the significant challenge of integrating quite different methods and explanations, such as experimental and theoretical approaches, microevolutionary and macroevolutionary models, developmental and population genetic explanations” (Brigandt, 2010: 298).

⁶In 1977, Darden and Maull presented *interlevel* theories as a “subset” of interfield theories” (Maull, 1977: 160). Yet, they were more concerned with the relationships between the different methods, techniques, explanatory goals, and vocabularies of different ‘areas of research’ for solving some focal problems at different ‘levels of description’ than with theories of phenomena located at different “structural levels”, i.e., in terms of ‘part/whole’ orderings or relationships (Maull, 1977: 154; Darden & Maull, 1977: 44).

⁷According to Lévy, reduction by synthesis often results “in the construction of intermediary disciplines”. Therefore, “biology can now follow the example of chemistry in constructing those *intermediary* domains characteristic of reduction, without fearing mutilation; for a discipline is *enriched* when approached by the method of ‘reduction by synthesis’” (Lévy, 1980: 157 – italics inserted). For a recent account of Lévy’s notion of ‘reduction by synthesis’, see Bensaude-Vincent and Simon (2012: 164–168).

The set of problems and questions raised by the different forms of integration is so vast and complex that it cannot be discussed here. However, in the context of our analysis, the most important point issue to emphasize is that it is the objective existence of processes of systemic emergence and downward determination (causal or otherwise) that most strongly necessitates the use of interlevel integrative forms of explanation.

In order to explain emergent systems' properties and downward causal processes, we need to do more than make higher-level and lower-level descriptions compatible. We need an adequate articulation of upward and downward *explanations*. In other words, we need to explain (i) *why* lower-level systems behave the way they do as constitutive *relata* of some higher-level systems of relations, and (ii) *how* some system-level properties result from specifically organized systems of relations between their constitutive *relata* as lower-level parts. In sum, the critical issue is to explain how levels of organization *relate* to each other, *partially determining* each other's existence and identity.

12.4.2 *Inter-theoretical Relations*

We can also address this issue from the point of view of inter-theoretical relations. For example, under what conditions can a theory of a system *w*, of a kind *K*, be literally reduced to the theory about the lower-level individuals (*Ls*), of a kind *L*, that are or may be proper parts of *w*?

Can the theory of the intrinsic or *w*-independent properties, and respective causal and nomological relations, of *Ls*, fully explain all non-relational properties of system *w*? Only in this case, we could talk about a proper reduction of a *higher-level theory* to a *lower-level theory*. If some system properties of *w* are only completely determined or produced by (thereby only being fully explainable in terms of) a specific organization between some properties which *Ls* only acquired or manifest as parts of or *relata* within the specific system of relations which obtain in *w*, then in no meaningful way we can talk about a literal lower-level reduction. The fact that the complete explanation of *w*'s properties requires the *incorporation* of some *w*-dependent relational properties of *Ls* just means that the lower level is *not*, in itself and by itself, *enough* to account for all the higher-level properties of the system *w*.

Imagine that one could build a general theory (*G*) incorporating *all* properties that *Ls* can acquire or manifest in *every* possible relational context, including as proper parts of systems of the kind *K*. Then, if we abstract away from all exogenous relations that such systems may have with other systems (including in the context of further, higher-level systems of relations), we might say that all non-relational properties of systems of the kind *K* can be fully explained in terms of theory *G*. But what kind of explanation would that be? Could theory *G* be considered a literal lower-level theory, thereby enabling a proper lower-level reduction of the higher-level theory of systems of kind *K*? The answer can only be negative. Once a lower-level theory begins to incorporate all relational properties and behaviors that its systems acquire or manifest in *all* possible relational contexts, including as proper parts of higher-level systems,

it ceases to be a pure lower-level theory. That theory is already a *new* theory, changed and enriched by the integration of all higher-level relevant factors.

The theory-reduction model faces the same challenges as any mechanistic explanation that opts for a more ‘localized’ or ‘particularist’ approach.

Even if someone were to defend that all such system-dependent relational properties or causal powers were already possessed by the lower-level systems as *intrinsic* dispositions (as contemporary individualistic essentialists would argue), that would still not allow for proper lower-level reductive explanations. As I have argued elsewhere (Santos, 2020: 8695–8696; also: 2015a, 2021), even if the intrinsic dispositions thesis were right, one would still need to explain why some potentialities are actualized in certain relational settings but *not* in others, why some potentialities are actualized *instead* of others (including their direct opposites), and why some potentialities are not even *ever* actualized. To account for all this, we need relations and systems of relations as absolutely necessary *ontological* and *explanatory* factors. Indeed, without relations, including second-order and even higher-order ones, the alleged intrinsic potentialities of all lower-level entities would remain latent and inactive for all eternity. That is, they would never come into actual existence, thereby failing to have the causal effects that they actually have on the dynamics and structure of our universe.

Consider cell types. Cells differentiate depending on the properties of their developmental contexts. This means that if the same cells were put in a different tissue, they would differentiate differently (Vecchi, 2020b: 62–63). As Soto and Sonnenschein observed, “[a] single cell isolated from either one of these tissues (...) fails to originate the tissues that would result from their reciprocal interactions” (2011: 333). A particularly significant instance of RTE and DC is that when cancerous cells are transplanted or injected into healthy tissues, their behavior is ‘normalized’, reverting to a non-cancerous state (Soto & Sonnenschein, 2011: 338). How lower-level would then be a theory of cells integrating all properties that cells may acquire by virtue of being parts of or related within *higher-level* organized systems of relations, such as *tissues* or *organisms*?

This problem, of course, is not new. According to Robert Causey, a lower-level reduction of a higher-level theory may be obtained if scientists “study the behavior of the components of structured wholes when they are *not* part of the whole (...) and then derive their behavior when part of the structured whole from this information *plus* specification of the boundary conditions prevailing when they are bound” (Bechtel & Hamilton, 2007: 398). Cliff Hooker and Patricia Churchland followed another, but similar strategy: “to incorporate into the lower-level theory everything that is learned about lower-level entities as they are bound into various structured wholes” (Bechtel & Hamilton, 2007: 398). In sum, “lower-level theories need to be enriched to account for what is learned at the higher-level” (Bechtel & Hamilton, 2007: 399).

As a matter of fact, neo-positivists were also well aware of the necessity of *changing* the lower-level theories by *enriching* them with all the necessary higher-level factors. For example, Nagel acknowledged that some systems (which he called ‘organic’ or ‘functional’) cannot be fully explained by and thus reduced to the laws relating the properties which their proper parts manifest *independently* of being parts of those systems. In such systems, parts “stand in relations of causal

interdependence”, that is, they “do not act, and do not possess characteristics, *independently* of one another” (Nagel, 1961: 395, 391). Therefore, “any laws which may hold for such parts when they are not members of a functional whole cannot be assumed to hold for them when they actually are members” (1961: 394). Any additive analysis of organic or functional wholes “must *include* special assumptions about the *actual organization* of parts in those wholes when it attempts to apply some fundamental theory to them”. In sum, the explanation of such systems “in terms of theories about their constituent parts cannot avoid *supplementing* these theories with statements about the special circumstances under which the constituents occur as elements in the systems” (1961: 395 – italics added).

Hempel also recognized the need to *supplement* the lower-level theories with information relative to specific higher-level systems of relations. In particular, Hempel argued that the complete explanation of some wholes could be provided only if in addition to the independent properties of their parts, we integrate all “relational information” concerning the “spatial or other relations”, including “structural relations”, among the parts (Hempel, 1965: 260–261). A complete explanation would then require a “description, in terms of relational concepts, of the way in which [parts] are connected with each other” in each different kind of whole (1965: 261).⁸

The problem with all these strategies is not so much the real possibility of building such general theories as the *epistemological meaning* of that possibility.

Consider, again, a general theory (*G*) that incorporates *all* properties that some entities (*Ls*) can acquire or manifest in *all* possible relational contexts, including as proper parts of systems of the kind *K*. If we took those systems as *isolated* and, for the sake of argument, we also ignored the historical processes that lead to the formation of their organizational structures, we could, of course, deliver a complete explanation of such system in terms of a *G* theory about *Ls*. The question is that *that* explanation would no longer represent a proper lower-level reduction. That theory would just be an example of an *interlevel integrative theory*, as the epistemological expression of the ontological reciprocal determination between lower and higher levels of organization.

The theory-reduction model attempted to assimilate that reciprocal determination by explicitly invoking the need of adding specific *boundary conditions* when reducing higher- to lower-level laws. The problem is that boundary conditions “are

⁸Hempel and Nagel were, of course, aware of the criticisms elaborated by their contemporary organismic views in biology, as well as by some holistic views in psychology (e.g., Gestalt theory) and in physics (e.g., the field theories). Bertalanffy, for example, criticized the “analytical, summative and machine theoretical viewpoints” of modern science, whose ultimate goal was “to explain phenomena by reducing them to an interplay of elementary units that could be investigated *independently* of each other” or “to resolve all natural phenomena into a play of elementary units, the characteristics of which remain *unaltered* whether they are investigated in *isolation* or in a *complex*”. In his view, the “organismic conceptions assert the necessity of investigating not only parts but also *relations of organisation* resulting from dynamic interaction and manifesting themselves by the *difference* in behavior of parts in isolation and in the whole organism” (1950: 134–135 – italics inserted).

not themselves derived from the lower-level laws”. Additionally, “where do these boundary conditions come from?” (Bechtel & Hamilton, 2007: 399; see also: Bishop et al. 2022: 278–283).⁹

Some systems may be affected by same-level or even lower-level boundary conditions, but some boundary conditions are clearly the result of higher-level organizations (Noble et al., 2019). In the latter case, as Bechtel notes, “by just characterizing such information as specifying boundary conditions and not considering what that information is about, namely, the organization involved in constituting a *higher-level system* out of lower-level constituents, the theory-reduction account *camouflages* the contribution of higher-level inquiries” (Bechtel, 2007: 150, n. 6 – italics inserted).

Specific modes of organization are typically relegated to the status of boundary conditions by reductionists. Yet, as Bechtel observed,

insofar as the boundary conditions cohere into *stable structures* that are *heritable*, they acquire a significant status and must be accommodated in any general endeavor to describe the course of events. After they arise, some of these stable structures may be *perpetuated*, and “once it is recognized that these *organizational structures* are the result of an *historical process*, the significance of any attempt to give a reductionistic explanation is radically reduced. To complete the reduction, one must fill in the details of the boundary conditions as they have historically arisen, a task that *cannot* be completed with just the *laws of the basic theory* (Bechtel, 1986b: 97 – italics inserted).

Furthermore, according to Bishop (2019), in addition to the boundary conditions, we must also take into account the existence of *stability conditions*, which “don’t function like boundary conditions”, but rather constitute necessary conditions for the very “existence and persistence of appropriate states and observables and systems” (2019: 5.7). In fact, without specific stability conditions, there wouldn’t even be nothing to which laws and boundary conditions could apply. And yet stability conditions are “never given by”, nor they are “derivable from the underlying scale or domain alone” (Bishop, 2019: 3.2; also: Bishop et al. 2022: 27–36 and 275–278).¹⁰

⁹As Jean Ullmo has long ago observed, “[e]very law is relative to an isolated system, that is, it describes a specific type of interaction in which an object may be involved, abstraction made of all interactions that take place simultaneously, and which one assumes, or makes sure to be, negligible” (1958: 156). In this sense, all “repeatable relations” expressed by law-statements are “conditioned”, and “it is [their] implicit conditions that ensure their validity” (1958: 54). The notion of scientific laws as exceptionless universal statements with an unlimited scope is a scientifically unfounded philosophical myth. Unfortunately, this caricatured interpretation of laws has often been used to minimize or even call into question the existence of laws in nature and the role they play in the construction of scientific knowledge.

¹⁰According to Bishop, Silberstein and Pexton, “Not only can stability conditions arise dynamically (and locally) constraining the behavior of basic components (e.g., convection), or be causal-mechanical constrains such as mechanical equilibrium, but stability conditions also include global or systemic constrains such as topological constrains, dimensional constrains, network or graphical constrains, and order parameters, among others. One can point to both dynamical and adynamical, causal and acausal, as well as local and global stability conditions that are difference makers. For example, global adynamical constrains might include conservation laws, free energy principles, least action principles, symmetries, and some types of symmetry breaking” (2022: 28).

All of this shows the reason why the causal closure or completeness principle of any level of organization should be discarded as simply wrong (Bishop et al., 2022: 288–303).

Another related issue (already mentioned) comes with the fact that, very often, the accounts that present themselves as constituting lower-level explanations are grounding their explanations in lower-level individuals taken *as already determined* parts of or related within specific higher-level systems. Sarkar (2015) points to this problem in Nagel’s model of inter-theoretical reduction. Reductions often involve approximations and, in particular, approximations on which the derivation of higher-level from lower-level theories depend. Still, there are approximations that “are justified by the reducing theory”, while others “are tailored to fit the reduction and implicitly rely on the reduced theory”. As Sarkar notes, “the serious question [is] whether a reduction *only* invokes as *explanans* factors that are indubitably from the reducing theory” (2015: 50). In sum, the problem is that of *presupposing* higher-level determining factors when we elaborate our lower-level theories.

As Bishop, Silberstein and Pexton note, many of the “alleged” inter-theoretic reductions require to “implicitly import some of the wider contextual features at the higher-level without acknowledging them” (2022: 69).

This problem is debated in most if not all sciences. Consider a prototypical example of a holist explanation in the social sciences: “the rise in unemployment led to a higher crime rate”. Now consider the alternative individualist explanation: “as a result of individuals *a, b, c*, etc. losing their job and feeling very frustrated about having little money and no job opportunities, the crime rate went up” (Zahle & Kincaid, 2019: 659). The relevant question in this context is *how micro- or low-level* is an explanation of a social phenomenon that adduces in its *explanans* facts such as ‘having little money’ and ‘having no job opportunities’. These two properties are clearly not psychological. Both are extrinsic relational properties that individuals acquire solely by virtue of living in a specifically organized structure of socio-economic relations. Indeed, some properties combine properties pertaining to different levels of organization. Think of a complex property such as ‘being afraid of losing her job’. What would constitute a real explanation of the instantiation of that complex property? The first property is clearly psychological, referring to a particular mental state, while the second is a clear socio-economic extrinsic relational property. As Auyang pointed out, “[a]lmost all explanations in terms of concrete individuals involve social predicates. Thus, social concepts have not been eliminated as demanded by methodological individualism; *they have only been swept under the rug*” (1999: 357, n.1 – italics inserted). As Kincaid has noted, the problem is that of “presupposing the reduced theory in the reducing explanations” (2012: 148), that being the reason why “many so called individualist explanations are really individualist only in name” (2012: 149).¹¹

¹¹ Margaret Archer (1995)’s morphogenetic social theory is an attempt to overcome both individualistic and holistic approaches in contemporary sociological thinking.

Nevertheless, from the vantage point of a relational view, there is no reason to see a conflict between higher-level explanations (e.g., sociological) in terms of some wholes (e.g., systems of social relations) and lower-level explanations (e.g., psychological) in terms of their proper parts (e.g., human beings), for they “complement each other in revealing the dual aspect, individual and inter-individual, of all behaviour patterns in human society” (Piaget, 1995: 41).

That complementarity is clearly exemplified in the set of all *system-dependent relational properties* that entities acquire and manifest as *relata* of specifically organized systems of interdependent transformative relations. Just as each human being is an already *socialized* individual when living in a specifically organized system of social relations, each atom is an already *molecularized* atom when part of a specifically organized molecular system of interdependent relations, and each molecule (such as DNA) is an already *cellularized* molecule when part of a specifically organized cellular system of interdependent relations. This is how we should conceptualize individual entities as *relata* within integrated relational systems. For example, a human being is not only a bio-neuro-psychological system but a *socio-economically* and *culturally shaped* bio-neuro-psychological being (e.g., Archer, 1995; Bishop et al., 2022: 223–226 and Lawson, 2019).¹²

A higher-level organization cannot be just anything, that is, regardless of the entities composing it (one cannot build living cells with crystals). That is why there is always a partial micro-determination of the higher levels. But the lower-level properties and laws are not enough to completely determine, and thus to fully explain, not just all higher-level systems but also the behaviors of the lower-level entities as parts of such systems. In sum,

The arrow of determination and explanation is not exclusively bottom-up but multi-scale and multidirectional, since any causal process is bounded by relational constraints which can be top-down, bottom-up, or side to side (as it were). There are no discrete causally closed or absolutely autonomous scales or domains of reality. Rather, there is a relation of mutual integration, interdependence, and reciprocal conditioning (Bishop et al., 2022: 283).

This is the reason why in many, if not in all cases, “the higher-level theories (for instance, cell physiology) and the lower-level theories (for instance, biochemistry) are ontologically and epistemologically inter-dependent on matters of informational content and evidential relevance” (Cat, 2022). In other words, scientific explanations often require “a genuine ‘*hybridization*’, with fruitful re-combinations”, between different disciplines or domains of research, where “the link between a ‘higher’ (in the sense of ‘more complex’) and a ‘lower’ field results neither in a reduction of the first to the second nor in greater heterogeneity of the first, but in *mutual assimilation* such that the second explains the first, but does so by *enriching* itself with *properties* not previously perceived, which afford the *necessary link*” (Piaget, 1970c: 525 – italics inserted).

¹²In this regard, it is worth mentioning that the causal impact of socioeconomic factors on brain structure and functioning, as well as cognitive and emotional development, is a growing research topic in current *social neuroscience* (e.g., Troller-Renfree et al., 2022; Thomas & Coecke, 2023).

This process of enrichment/reciprocal assimilation is paradigmatic in cases of real interlevel integration. For example, according to biophysicist Henri Atlan, it is not true that ‘life’ has been literally reduced to physical chemistry. What happened is that physical chemistry was changed and “extended”, thereby allowing the creation of a “biophysics of organized systems” (Atlan, 1979: 23–24). Similarly, biochemistry is not just ‘applied chemistry’, since it already constitutes “an *extension* relative to mineral and organic chemistry” (Atlan, 1979: 24, n.1; see also Bechtel, 1986b). And the same could be said about the relationships between quantum mechanics and chemistry, since the laws of the former do not fully determine quantum chemistry (Bishop, 2019, sections 4.13–4.18).

The process of *enriching* lower-level theories, as well as the construction of intermediate theories between lower- and higher-level theories, is just the *epistemological* replication of the ontological process of *transformation* that lower-level entities undergo in terms of the *relational* properties they acquire and manifest as relata within higher-level structured systems of relations.

12.4.3 *Some Implications for a Neo-mechanistic Model of Explanation*

This process of enrichment/reciprocal assimilation is also evident in any neo-mechanistic model of explanation. At least in the case of integrated systems, *reduction* only refers to the necessary, but in itself insufficient, *analytical* methodological step of the explanation process concerning the operations of decomposition and localization. First, because these operations must be followed by two other methodological steps, namely, the *synthetic* operations of ‘recomposing’ and ‘situating’ a mechanism as a whole (see Sect. 12.2). Second, because these last two operations often show that the previous decompositions and localizations must be revised and corrected (Bechtel & Richardson, 2010: xxxvii–xl, *et passim*). Therefore, the explanatory tasks of recomposing and situating determine the very operations of decomposition and localization by showing the ways in which these may (and may not) be carried out (Bechtel, 2002; Bechtel & Abrahamsen, 2010; Bishop et al., 2022: 235–239).

Two conclusions can be drawn from this. First, the methodological steps of a neo-mechanistic model of explanation do not progress themselves in a linear, sequential fashion, but rather constitute a cyclic process with feedback consequences. Second, the explanatory task of ‘situating’ should be applied not only to a system or mechanism as a whole, but *also* to its *parts* since parts are also dependent both on intra- and extra-systemic relations. As it was argued, the properties and laws of any class of entities, taken as independent beings or isolated systems, only provide the necessary but not the sufficient conditions, not just to ontologically determine any higher-level organization but also to determine their own existence and identity.

Acknowledgements I acknowledge the financial support of FCT, ‘Fundação para a Ciência e a Tecnologia, I.P.’ (Stimulus of Scientific Employment, Individual Support 2017: CEECIND/03316/2017). Thanks to Davide Vecchi for his helpful comments and suggestions. Thanks to an anonymous reviewer for her/his stimulating feedback.

References

- Archer, M. (1995). *Realist social theory: The morphogenetic approach*. Cambridge University Press.
- Atlan, H. (1979). *Entre le cristal et la fumée. Essai sur l’organisation du vivant*. Seuil.
- Atlan, H. (2011a). *Le Vivant post-génomique. Ou qu’est-ce ce l’auto-organisation?* Odile Jacob.
- Atlan, H. (2011b). Noise as a principle of self-organization. In S. Geroulanos & T. Meyers (Eds.), *Henri Atlan. Selected writings. On self-organization, philosophy, bioethics, and Judaism* (pp. 95–113). Fordham University Press.
- Atlan, H., & Cohen, I. R. (2006). Self-organization and meaning in immunology. In B. Feltz, M. Crommelinck, & P. Goujon (Eds.), *Self-organization and emergence in life sciences* (pp. 121–139). Springer.
- Atlan, H., & Koppel, M. (1990). The cellular computer DNA: Program or data? *Bulletin of Mathematical Biology*, 52(3), 335–348.
- Auyang, S. (1999). *Foundations of complex-system theories – In economics, evolutionary biology, and statistical physics*. Cambridge University Press.
- Bechtel, W. (1986a). The nature of scientific integration. In W. Bechtel (Ed.), *Integrating scientific disciplines* (pp. 3–52). Martinus Nijhoff.
- Bechtel, W. (1986b). Biochemistry: A cross-disciplinary endeavor that discovered a distinctive domain. In W. Bechtel (Ed.), *Integrating scientific disciplines* (pp. 77–100). Martinus Nijhoff.
- Bechtel, W. (2002). Decomposing the brain: A long-term pursuit. *Brain and Mind*, 3, 229–242.
- Bechtel, W. (2006). *Discovering cell mechanisms: The creation of modern cell biology*. Cambridge University Press.
- Bechtel, W. (2007). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge.
- Bechtel, W. (2011). Mechanism and biological explanation. *Philosophy of Science*, 78(4), 533–557.
- Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science*, 41, 321–333.
- Bechtel, W., & Hamilton, A. (2007). Reduction, integration, and the unity of science: Natural, behavioral, and social sciences and the humanities. In T. A. F. Kuipers (Ed.), *Philosophy of science: Focal issues* (pp. 377–430). Elsevier.
- Bechtel, W., & Richardson, R. (2010). *Discovering complexity: Decomposition and localization as strategies in scientific research*. The MIT Press.
- Bensaude-Vincent, B., & Simon, J. (2012). *Chemistry: The impure science* (pp. 164–168). Imperial College Press.
- Bertalanffy, L. (1950). An outline of general system theory. *British Journal for the Philosophy of Science*, 1, 134–165.
- Bishop, R. C. (2008). Downward causation in fluid convection. *Synthese*, 160, 229–248.
- Bishop, R. C. (2019). *The physics of emergence* (IOP concise physics series). Morgan & Claypool Publishers.
- Bishop, R. C., & Atmanspacher, H. (2006). Contextual emergence in the description of properties. *Foundations of Physics*, 36, 1753–1777.

- Bishop, R. C., Silberstein, M., & Pexton, M. (2022). *Emergence in context. A treatise in twenty-first century natural philosophy*. Oxford University Press.
- Brigandt, I. (2010). Beyond reduction and pluralism: Toward an epistemology of explanatory integration in biology. *Erkenntnis*, 73, 295–311.
- Brigandt, I. (2013). Integration in biology: Philosophical perspectives on the dynamics of interdisciplinarity. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 461–465.
- Bunge, M. (1959). *Causality. The place of the causal principle in modern science*. Harvard University Press.
- Bunge, M. (1964). Phenomenological theories. In M. Bunge (Ed.), *The critical approach to science and philosophy* (pp. 234–254). Free Press.
- Bunge, M. (1967). *Scientific research II: The search for truth*. Springer.
- Bunge, M. (1982). The revival of causality. In G. Fløistad (Ed.), *La philosophie contemporaine / contemporary philosophy: Chroniques nouvelles / A new survey* (Vol. 2, pp. 133–155). Springer. https://doi.org/10.1007/978-94-010-9940-0_6
- Bunge. (1983). *Treatise on basic philosophy* (Vol. 6: Epistemology & methodology II). D. Reidel.
- Cat, J. (2022). The unity of science. In: E.N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Spring 2022 Edition). <https://plato.stanford.edu/archives/spr2022/entries/scientific-unity/>
- Cohen, I. R., Atlan, H., & Efroni, S. (2016). Genetics as explanation: Limits to the human genome project. In *Encyclopedia of life sciences*. Wiley. <https://doi.org/10.1002/9780470015902.a0005881.pub3>
- Craver, C. (2005). Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 36, 373–395.
- Craver, C., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, 22(4), 547–563.
- Craver, C., & Darden, L. (2013). *In search of mechanisms: Discoveries across the life sciences*. University of Chicago Press.
- Craver, C., & Tabery, J. (2019). Mechanisms in Science. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Summer 2019 Edition). <https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/>
- Darden, L., & Maull, N. (1977). Interfield theories. *Philosophy of Science*, 44, 43–64.
- Donati, P., & Archer, M. (2015). *The relational subject*. Cambridge University Press.
- Gilbert, S. F. (2010). *Developmental biology* (9th ed.). Sinauer Associates.
- Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. The Free Press/Collier-Macmillan Ltd.
- Hilgetag C. C. and Goulas, A. (2020), ‘Hierarchy’ in the organization of brain networks”, *Philosophical Transactions of the Royal Society B: Biological Sciences* 375(1796): 20190319. doi:<https://doi.org/10.1098/rstb.2019.0319>.
- Hooker, C. (2013). On the import of constraints in complex dynamical systems. *Foundations of Science*, 187, 757–780.
- Humphreys, P. (2016). *Emergence. A philosophical account*. Oxford University Press.
- Keller, E. (1995). *Refiguring life. Metaphors of twentieth-century biology*. Columbia University Press.
- Keller, E. (2001). Beyond the gene but beneath the skin. In S. Oyama, P. Griffiths, & R. Gray (Eds.), *Cycles of contingency: Developmental systems and evolution* (pp. 299–312). MIT Press.
- Keller, E. (2010). It is possible to reduce biological explanations in chemistry and/or physics? In F. Ayala & R. Arp (Eds.), *Contemporary debates in philosophy of biology* (pp. 19–31). Wiley-Blackwell.

- Kincaid, H. (2012). Some issues concerning the nature of economic explanation. In U. Mäki, D. Gabbay, P. Thagard, & J. Woods (Eds.), *Handbook for the philosophy of science: Philosophy of economics* (pp. 137–159). Elsevier.
- Klee, R. (1984). Micro-determinism and concepts of emergence. *Philosophy of Science*, 51, 44–63.
- Lawson, T. (2013). Emergence and social causation. In R. Groff & G. Greco (Eds.), *Powers and capacities in philosophy: The new Aristotelianism* (pp. 61–84). Routledge.
- Lawson, T. (2016). Some critical issues in social ontology: Reply to John Searle. *Journal for the Theory of Social Behaviour*, 64(4), 426–437.
- Lawson, T. (2019). *The nature of social reality. Issues in social ontology*. Routledge.
- Levins, R. (1970). Complex systems. In C. H. Waddington (Ed.), *Organization, stability & process: Toward a theoretical biology* (Vol. 3, pp. 73–87). Routledge.
- Lévy, M. (1979). Les relations entre chimie et physique et le problème de la réduction. *Epistemologia*, 2, 337–370.
- Lévy, M. (1980). The ‘reduction by synthesis’ of biology to physical chemistry. In *PSA: Proceedings of the biennial meeting of the philosophy of science association* (Vol. 1: Contributed papers) (pp. 151–159).
- Lewontin, R., & Levins, R. (2007). *Biology under the influence: Dialectical essays on ecology, agriculture, and health*. Monthly Review Press.
- Love, A. C., & Lugar, G. L. (2013). Dimensions of integration in interdisciplinary explanations of the origin of evolutionary novelty. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 537–550.
- Maull, N. (1977). Unifying science without reduction. *Studies in History and Philosophy of Science*, 8, 143–162.
- Nagel, E. (1961). *Structure of science: Problems in the logic and scientific explanation*. Harcourt, Brace & World.
- Noble, R., Tasaki, K., Noble, P. J., & Noble, D. (2019). Biological relativity requires circular causality but not symmetry of causation: So, where, what and when are the boundaries? *Frontiers in Physiology*, 10, 827. <https://www.frontiersin.org/articles/10.3389/fphys.2019.00827>
- Piaget, J. (1950a). Introduction à l'épistémologie génétique. vol. 2: *La pensée physique*. PUF.
- Piaget, J. (1950b). Introduction à l'épistémologie génétique. vol. 3: *La pensée biologique, la pensée psychologique et la pensée sociale*. PUF.
- Piaget, J. (1963). L'explication en psychologie et le parallélisme psychophysologique. In P. Fraisse & J. Piaget (Eds.), *Traité de psychologie expérimentale* (Vol. I, pp. 137–184). PUF.
- Piaget, J. (1967a). Les relations entre le sujet et l'objet dans la connaissance physique. In J. Piaget (org.), *Logique et Connaissance Scientifique* (pp. 754–778). Gallimard.
- Piaget, J. (1967b). Classification des sciences et principaux courants épistémologiques contemporains. In J. Piaget (org.), *Logique et Connaissance Scientifique* (pp. 1151–1271). Gallimard.
- Piaget, J. (1970a). Introduction: The place of the sciences of man in the system of sciences. In *Main trends of research in the social and human sciences, part 1: Social sciences* (pp. 1–57). Mouton/UNESCO.
- Piaget, J. (1970b). Psychology. In *Main trends of research in the social and human sciences. Part one: Social sciences* (pp. 225–282). Mouton/UNESCO.
- Piaget, J. (1970c). General problems of interdisciplinary research and common mechanisms. In *Main trends of research in the social and human sciences. Part one: Social sciences* (pp. 467–528). Mouton/UNESCO.
- Piaget, J. (1971a). *Biology and knowledge. An essay on the relations between organic regulations and cognitive processes*. The University of Chicago Press.
- Piaget, J. (1971b). *Structuralism*. Routledge & Kegan Paul.
- Piaget, J. (1995). *Sociological Studies*. Routledge.

- Povich, M., & Craver, C. (2018). Mechanistic levels, reduction, and emergence. In S. Glennan & P. Illari (Eds.), *The Routledge handbook of mechanisms and mechanical philosophy* (pp. 185–197). Routledge.
- Salmon, W. (1989). Four decades of scientific explanation. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (Minnesota studies in the philosophy of science) (Vol. 13, p. 3). University of Minnesota Press.
- Santos, G. (2015a). Ontological emergence: How is that possible? Towards a new relational ontology. *Foundations of Science*, 20(4), 429–446.
- Santos, G. (2015b). Upward and downward causation from a relational-horizontal ontological perspective. *Axiomathes*, 25(1), 23–40.
- Santos, G. (2020). Integrated-structure emergence and its mechanistic explanation. *Synthese*, 198, 8687–8711. <https://doi.org/10.1007/s11229-020-02594-3>
- Santos, G. (2021). Emergentism. In V. P. Glăveanu (Ed.), *The Palgrave encyclopedia of the possible* (pp. 1–8). Palgrave Macmillan. https://doi.org/10.1007/978-3-319-98390-5_167-1
- Santos, G., Vallejos, G., & Vecchi, D. (2020). A relational-constructionist account of protein macrostructure and function. *Foundations of Chemistry*, 22, 363–382.
- Sarkar, S. (2015). Nagel on reduction. *Studies in History and Philosophy of Science*, 53, 43–56.
- Shapiro, J. (2009). Revisiting the central dogma in the 21st century. *Annals of the New York Academy of Sciences*, 1178, 6–28.
- Simons, P. (1987). *Parts. A study in ontology*. Clarendon Press of Oxford University Press.
- Simons, P. (2006). Real wholes, real parts: Mereology without algebra. *The Journal of Philosophy*, 103, 597–613.
- Soto, A. M., & Sonnenschein, C. (2011). The tissue organization field theory of cancer: A testable replacement for the somatic mutation theory. *BioEssays*, 33, 332–340.
- Strohman, R. (1997). Epigenesis and complexity: The coming Kuhnian revolution in biology. *Nature Biotechnology*, 15, 194–200.
- Thomas, M. S. C., & Coecke, S. (2023). Associations between socioeconomic status, cognition and brain structure: Evaluating potential causal pathways through mechanistic models of development. *Cognitive Science*, 47, e13217. <https://doi.org/10.1111/cogs.13217>
- Troller-Renfree, S. V., Costanzo, M. A., Duncan, G. J., Magnuson, K., Gennetian, L. A., Yoshikawa, H., Halpern-Meekin, S., Fox, N. A., & Noble, K. G. (2022). The impact of a poverty reduction intervention on infant brain activity. *Proceedings of the National Academy of Sciences of the USA*, 119(5), e2115649119. <https://doi.org/10.1073/pnas.2115649119>
- Ullmo, J. (1958). *La Pensée Scientifique Moderne*. Flammarion.
- Vecchi, D. (2020a). DNA is not an ontologically distinctive developmental cause. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 81. <https://doi.org/10.1016/j.shpsc.2019.101245>
- Vecchi, D. (2020b). Organismality grounds species collective responsibility. *Rivista di estetica*, 75(3), 52–71.
- Vecchi, D., & Santos, G. (2023). The multi-causal basis of developmental potential construction. *Acta Biotheoretica*, 71(6). <https://doi.org/10.1007/s10441-023-09456-8>
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
- Wiener, N. (2019). *Cybernetics or control and communication in the animal and the machine*. The MIT Press.
- Wimsatt, W. (1997). Aggregativity: Reductive heuristics for finding emergence. *Philosophy of Science*, 64(4), S372–S384.
- Wimsatt, W. (2000). Emergence as non-Aggregativity and the biases of reductionism(s). *Foundations of Science*, 5, 269–297.
- Wimsatt, W. (2006). Aggregate, composed, and evolved systems: Reductionistic heuristics as means to more holistic theories. *Biology and Philosophy*, 21, 667–702.
- Wimsatt, W. (2018). Foreword. In S. Glennan & P. Illari (Eds.), *Routledge handbook of mechanisms and mechanical philosophy* (pp. xiv–xvi). Routledge.
- Zahle, J., & Kincaid, H. (2019). Why be a methodological individualist? *Synthese*, 196, 655–675.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Index

A

Abstraction, 94, 140, 141, 144, 236, 245, 247
Action at a distance, 183
Activities, 11–13, 16–18, 21, 23, 25, 30, 31, 33, 34, 36, 39, 41, 51, 52, 58, 62, 65, 67, 69, 71, 72, 74, 76, 78, 79, 86–88, 90–94, 96–98, 100–105, 112, 118, 119, 125–127, 131, 144, 146, 147, 192, 201, 202, 207, 215–221, 223–225, 228, 229, 231, 235, 239, 245, 247
Actual identity, 237, 241
Adenosine triphosphate (ATP), 18, 90–92, 113
Adiabatic, 146, 154–158
Aggregation, 32, 114, 223
Aggregativity, 222, 223
Allometry, 38, 40, 45, 46, 48, 54–57
Alphafold, 134
Approximations, 49, 154–156, 163, 164, 258
Artificial neural networks, 206
Atom, 114, 119, 140, 142, 143, 145, 147–151, 167–172, 174, 175, 181, 191, 197, 198, 201, 203–206, 228, 231, 259
Atomic weight, 142
Atomism, 193, 194, 231, 236, 238, 239, 248
Atomist/individualist essentialism, 3
Autonomy, 6, 22, 214, 217, 219, 220, 222–230, 232
Autonomy of constraints, 87, 92–94
Autonomy school (or tradition), 5, 86, 87, 92
Autopoiesis (or autopoietic), 87, 92

B

Behaving entity view of constitution, 17–19
Black box, 3, 21
Bohr model, 162

Born Oppenheimer approximation, 150, 155, 162, 163
Bottom up intervention, 19
Boundary conditions, 32–33, 156, 255–257
Brain and mind relation, 206, 207, 209
Brownian motion, 121–123, 131, 132
Bulk property, 152

C

Causal betweenness, 16, 17, 19
Causal explanation, 20, 54–56, 66, 67, 76, 121, 128, 130, 146, 165, 194, 195, 200, 251
Causalism (or causalist), 76
Causal laws, 200, 201, 203, 232
Causal power, 89, 144, 145, 152, 153, 155, 164, 165, 227, 230, 232, 240, 248, 255
Causal process, 63, 121, 133, 200, 201, 203, 218, 244, 254, 259
Causal relations, 3, 15, 22, 25, 36, 37, 50, 63, 180, 196, 197, 199, 207, 225, 238, 241, 242, 245, 246, 251
Causal structure, 33, 55, 193, 194, 196, 200
Causation, 6, 10, 12, 17, 18, 20, 24, 25, 31, 32, 36, 50, 54, 62, 63, 69, 76, 87, 92, 145, 164, 165, 200, 224, 237, 242, 243, 245, 246, 249
Causation/constitution distinction (or causal constitutive relations), 66, 67
Cause, 9, 12, 13, 15, 17, 20, 24, 25, 32, 42, 47, 48, 53–55, 62, 63, 65–69, 74, 76, 92, 95, 96, 118, 128, 130, 150, 165, 170, 181, 192, 193, 195–197, 200, 203, 219, 225, 232, 240, 245
Causing, 31, 35, 71, 73–79

- Character identity mechanisms, 71–73
 Charge, 88, 119, 140, 145, 146, 149, 150,
 152–155, 174, 202–206, 209, 214
 Chemical bonding, 162–175
 Chemical kinetics, 40, 125, 147
 Chemical reactions, 5, 39, 100, 123, 132,
 140–145, 147, 152
 Chemical species, 40, 140
 Chemical substance, 16, 140, 141, 143,
 144, 152
 Chicago mechanists, 2, 251
 Chirality, 149
 Closure (of constraints) (closure to efficient
 causation), 87, 92–95, 97, 105
 Closure (organizational), 93, 94
 Community, 4, 38–40, 42, 45, 46, 51, 52, 58,
 168, 203, 224
 Component, 11–16, 18, 19, 21, 23–25, 30–32,
 34–37, 39, 40, 45, 48, 49, 52, 53, 58,
 62, 64–66, 68–70, 72–74, 76–79,
 86–90, 92–94, 96–98, 104, 105,
 110–114, 119, 122, 127, 130, 132–134,
 163, 167, 180, 182, 187, 192, 193, 196,
 197, 200–209, 216, 217, 220, 223, 225,
 226, 228, 229, 231, 239, 242, 243, 247,
 249, 251, 255
 Compound substance, 142
 Conditioning relations, 235, 246
 Configurational Hamiltonian, 163, 166
 Conservation laws, 203, 204, 206, 208, 257
 Conserved quantity, 5, 140, 145–147, 152,
 200, 203, 204, 207–209
 Constitutive mechanistic explanation, 10–17,
 22–25, 130, 218
 Constitutive relevance, 11–13, 16, 17, 19, 33
 Constrains, 48, 52, 53, 76, 87, 90, 93, 95, 96,
 105, 152, 224, 225, 228
 Constructive interference, 170–172
 Continuity requirement, 74
 Control mechanisms, 87, 88, 92, 95–105, 248
 Cooperativity, 65, 120, 121, 125, 127
 Correlation, 40, 50, 54, 207, 236
 Covalent bonding, 167–169, 172, 228
 Cybernetics, 2, 250, 251
- D**
- Decoherence, 166, 186, 187, 207
 Decomposable, 31, 34–36, 238, 243
 Decomposition, 3, 22, 30, 34, 37, 67, 75, 88,
 94, 120, 125–127, 146, 194, 196, 198,
 214, 222, 260
 Deductive-nomological, 198
- Deductive-nomological model of
 explanation, 29, 30
 Degrees of freedom, 89, 111, 122,
 133, 241
 Denaturation, 114
 Dependence, 6, 43, 49–51, 54, 56, 93, 104,
 121, 156, 180, 214, 215, 217,
 219, 222, 223, 226, 227, 230, 238,
 247, 252
 Destructive interference, 170–173
 Developmental bias, 76, 77, 79
 Developmental mechanism, 65, 73–75, 77
 Developmental potential, 249, 250
 Developmental repatterning, 63, 73, 77–80
 Diachronic emergence, 215, 218, 219, 221,
 227, 230
 Dimensioned approach, 21
 Dimensioned mechanistic explanation,
 4, 19–25
 Dispositional explanation, 74, 79
 Downward causal conditioning, 241, 248
 Downward causal transformation, 241, 249
 Downward causation, 5, 6, 36, 163, 165, 215,
 224–225, 235–237, 240–250
 Dynamical autonomy, 226–227
 Dynamical behaviour, 128, 156–158
 Dynamic mechanistic explanation, 36, 52, 65
 Dynamic properties, 202–207, 209
- E**
- Ecology, 1, 2, 4, 5, 29–58, 224
 Ecosystem, 4, 34, 39, 40, 42, 44–46, 49,
 51–53, 57, 58, 224
 Electron delocalization, 169
 Eliminative induction, 152
 Emergence, 2, 5, 6, 17, 32, 33, 37, 67, 74, 78,
 111, 140, 153, 154, 157, 158, 162–175,
 187, 213–232, 240–250
 Emergence (varieties of), 217, 218, 220,
 223, 230
 Emergence of pattern, 221
 Emergent activities, 219
 Emergent entities (or systems), 219
 Emergent interactions, 219
 Emergent processes, 219, 220
 Emergent properties, 158, 214, 219, 220, 223,
 224, 230, 238, 240
 Emergent relations, 218–220
 Endogenous relations, 235, 236, 247
 Energetic view of chemical bonding, 169
 Energetics, 5, 39, 51, 58, 113, 115, 116, 118,
 120, 131–134, 168–170, 174, 175

Energy, 38–45, 49, 52, 57, 86, 87, 89–93, 96, 114–117, 120, 121, 123, 128, 140, 144–146, 155, 169, 170, 172, 173, 200, 203, 204, 206, 209

Energy bias, 120

Energy landscape, 113, 115, 117, 120, 121, 124, 130–133

Ensemble, 121, 122, 124, 125, 130, 132, 134, 150

Enthalpy, 122, 129

Entropy, 51, 120, 122, 129, 133, 203

Epistemic reduction, 22, 23

Epistemological reduction, 2, 163–166

Equilibrium explanation, 128

Etiological mechanistic explanation, 9, 15, 19, 20, 25, 218, 219

Evolutionary innovation, 75–79

Evolutionary mechanism, 63, 75–79

Evolutionary novelty, 63, 70, 73, 74, 77, 79, 80

Exogenous relations, 236, 240, 247, 254

Experimental method, 191, 196, 253

Externalism, 223–224

F

Far from equilibrium, 86

First order relation, 244, 248

Fitness, 40, 47, 48, 224

Folding dynamic problem (FDP), 116, 117

Folding intermediate, 126

Folding pathway, 5, 115, 117, 123, 126

Foldon, 125–127

Force, 47, 53, 71, 89–92, 128, 131, 154, 155, 165, 173, 182–185, 188, 193, 196–198, 201, 202, 204, 226

Free energy, 5, 51, 58, 86, 87, 89–93, 95, 112–114, 117, 120, 125, 130, 131, 133, 146, 257

Functional, 15, 16, 21, 32, 34–38, 42, 44, 48, 68, 69, 92, 93, 111, 112, 130, 148, 245, 247–250, 255, 256

Fundamental physics, 6, 180–188

Fusion, 227–230

G

Generalization, 4, 16, 30, 33, 45, 53–56, 124, 129, 157, 192, 208, 228

Generative atomism, 231

Genotype-phenotype map (or structure), 75, 78, 79

Gibbs free energy, 89, 112, 129, 132

Global energy minimum, 115

Gravity, 3, 183–185

H

Hamiltonian operators, 162, 165, 173

Hellmann-Feynman theorem, 154

Heterarchical (organization), 99, 102

Hierarchical (organization), 32, 34, 66, 99, 105

Hierarchy, 52, 69, 98, 99, 144

High-level, 145

Holism, 6, 36, 49, 214, 217, 219, 220, 222–230

Holistic view (or notion), 236, 242, 256

Homeostasis, 100, 101, 119, 128, 251

Homology, 5, 63, 70–73, 76, 79

Hydrophobic effect, 110, 130, 133

Hypothalamus, 101–103

I

Idealizations, 202, 207

Individualistic essentialism, 236

Induction, 195, 197

Input-output functionalist view, 15, 19

Input output mechanistic explanation, 4, 19–21, 25

Inter-level causation, 36

Interlevel emergence, 187

Interlevel integrative explanations, 237, 239, 240, 250–260

Interlevel integrative theory, 256

Internal environment, 100

Interpretations of quantum mechanics, 185

Intervention, 13, 32, 54, 55, 183, 225, 244, 249

Interventionism, 10, 13, 18

Interventionist, 19, 54, 55, 203

Intrinsic potentialities (or dispositions), 248

Intrinsic properties (or system-independent), 110, 113, 117, 119–122, 127, 130, 132–134, 236–240, 249

Invariance, 54, 55, 222

Ionic bonding, 167, 168

Isomers, 5, 141, 143, 155, 162, 166

Isotope, 142, 143

K

Kinetic approach, 5, 115–118, 120–127, 130, 132–135

Kinetic energy, 41, 172–174

L

Law, 4, 29, 30, 32, 33, 37, 39, 45, 46, 54, 61, 64, 89, 91, 129, 152, 164, 165, 182, 183, 186, 187, 192, 193, 196, 198, 200, 201, 204, 206, 209, 216, 223, 228, 231, 232, 237, 238, 243, 251, 255–257, 259, 260

Levinthal's paradox, 114–115

Limit of mechanistic explanations, 208

Limits, 1, 5, 36, 42, 74, 89, 101, 167, 191–209, 245, 247, 252

Lineage explanation, 74, 76

Local energy minimum, 117

Localization, 3, 30, 34, 35, 37, 128, 150, 155–158, 260

Local relations, 3

M

Macro-reduction, 237

Mass, 40–43, 52, 54, 56, 140, 145, 146, 152, 154, 183, 199, 202–204, 206, 214, 220, 236

Measurement, 95–98, 103, 105, 184, 185, 187, 203, 207

The mechanical analogy, 202

Mechanical philosophy, 6, 62, 180–186, 188, 230

Mechanism-dependence, 215, 217–222, 226

Mechanistic constitution, 10–12, 14–17, 19, 22, 225

Mechanistic explanation (types of), 2, 9–25

Mechanistic explanations, 2, 4, 5, 15, 30, 31, 33, 35, 37, 38, 45–47, 51, 55–58, 61, 62, 66, 67, 73, 74, 77, 118, 119, 129, 131, 180, 185, 191–209, 214, 224, 253, 255

Mechanistic level, 4, 16, 23, 24, 36, 39

Mechanistic ontology, 180–188

Mechanistic science, 62, 63, 76

Mechanistic systems, 215, 217, 218, 220

“Mechanistic turn”, 200, 230

Medieval Aristotelian tradition, 195

Medieval science and philosophy, 195

“Mental mechanism”, 207

Mereological sum, 152

Metabolic rate, 38–42, 44, 46, 47, 50–52, 56, 57

Metabolism, 4, 34, 38–45, 47–52, 57, 58, 65, 100, 220

Metastable state, 115

Method of analysis and synthesis, 6, 192, 194–197

Micro determinism, 6, 237

Micro-reduction (or lower-level reduction), 237, 238, 254–256

Microstructural essentialism, 141, 152

Microstructuralism, 141–144, 152

Minimal frustration, 131, 134

Model of explanation, 134, 252, 260

Models, 2, 21, 29, 30, 32, 35–37, 41, 45–48, 52–54, 57, 58, 64, 65, 67, 68, 72, 74, 76, 90, 114, 124, 129, 133, 145, 148, 155, 156, 165, 186, 191, 199, 203, 204, 206, 209, 229, 230, 250–254, 258

Modular, 35

Modularity, 32, 49

Molecular mechanisms, 71

Molecular orbital theory, 174, 175

Molecular population, 153

Molecular structure, 91, 145, 147, 153–158, 162–175

Morphogenetic mechanism, 71

Multilevel mechanism, 69

Multiple realization, 16, 22, 215, 226–227

Mutual dependency, 13, 14, 18, 19

Mutual manipulability, 13, 17, 18

N

Native state, 110, 111, 113–132

Native structure, 110–113, 115, 116, 120, 125, 128, 129, 247

Naturalised approach, 180

Natural selection, 38, 47, 48, 57, 58, 62, 67–69, 72, 76–79, 86, 129, 221

Negative feedback, 101, 102

Neo-positivism, 2, 3, 251

Neo vitalist view, 2

Neural mechanisms, 192, 199, 206–208

Neuromodulators, 103

New mechanisms, 1, 2, 5, 6, 94, 185, 186, 208, 250–254

Node, 48, 49, 248

Nomological theory of explanation, 2

Non-aggregativity (non-aggregative), 32, 214, 222–223, 243

Non-localizability, 185

Non-reductive physicalism, 16, 225, 226, 231

Novelty, 2, 5, 6, 73–75, 77, 78, 124, 154, 158, 214, 217, 219, 220, 222–232, 237, 250

Nuclear charge, 141–143

O

Old mechanism (old mechanist view), 2

Old mechanistic views, 250

Ontogenetic cause, 63

Ontological reduction, 22, 163–166, 205, 208
 Operation, 30–32, 34–37, 45, 52, 58, 70, 87, 88, 90, 91, 98, 119, 181, 201, 217, 221, 222, 239, 242, 260
 Organism-level mechanism (or cause), 68
 Organization, 4, 6, 17, 18, 24, 30–37, 39, 40, 42, 45, 47, 49, 50, 52, 57, 58, 62, 65–67, 69–72, 74, 78, 86, 87, 91–93, 99, 126, 130, 201, 214, 217, 226, 230, 238–240, 242–246, 251–254, 256–260
 Organizational emergence, 239
 Organizational view of homology, 72, 73
 Output mechanistic explanation, 4, 19, 20

P

Particles, 89, 157, 163, 165, 182, 184, 185, 192, 199, 201–205, 208, 228
 Part-whole relations, 3, 14, 15, 18, 196, 197, 199, 200, 202–205, 207, 208, 237, 241, 250
 Part whole relationship, 202
 Pathway, 5, 39, 65, 115, 117, 120–123, 125, 126, 131, 134, 135, 140, 150–152, 218
 Phenomenon, 2–4, 6, 9–20, 22–25, 30, 31, 33–35, 37, 38, 45, 49–52, 55, 56, 58, 61, 62, 65, 67–70, 76–78, 86, 88, 90, 91, 94, 114–116, 118–120, 125, 127, 129, 130, 133, 135, 168, 169, 180–183, 186, 187, 192–201, 203–208, 213–218, 224, 225, 229–232, 252, 253, 258
 Phenotypic variation, 70, 73, 76, 78
 Physical objects, 181, 182, 184, 187
 Physico-cellular mechanisms, 70
 Pluralistic ontology, 3
 Polypeptide chain, 110, 111, 113, 117, 119, 122, 123, 125, 130, 133, 247
 Population, 4, 34, 39, 40, 42, 43, 45, 46, 48, 49, 51, 52, 58, 62, 67, 68, 70, 75–78, 101, 117, 122–124, 134, 143, 221, 224, 253
 Population-level mechanism (or cause), 68, 69, 76, 78, 79
 Potential energy (PE), 146, 156, 172, 173
 Potential energy surface, 49, 146
 Potential identity, 238, 241
 Product structure, 140, 146, 148
 Production mechanisms, 87, 95–105
 Protein folding problem (PFP), 5, 110, 111, 114, 115, 133, 134
 Proximate cause(s), 62, 76

Q

Quantization, 163
 Quantum decoherence, 5, 6, 166
 Quantum entanglement, 3, 184
 Quantum mechanics (QM), 2, 5, 6, 147, 150, 153, 154, 156, 162, 163, 165, 175, 180, 184–188, 201–203, 205, 206, 208, 252, 260

R

Randomness, 54, 110, 114, 115, 121, 123–126, 131
 Reagent structure, 140, 152
 Realizer functionalist view, 16, 17, 19
 Reciprocal assimilation, 252, 253, 260
 Reciprocal causation, 62, 76
 Recomposition, 3, 37, 65, 67, 75
 Reduction by synthesis, 252, 253
 Reductionism, 37, 140, 152, 153, 164, 165, 213
 Reductionist (explanations), 55
 Reductive explanation, 4, 213, 214, 223, 226, 255
 Regression, 40, 41, 45, 52–54
 Relational, 6, 119, 134, 185, 187, 235–238, 240, 242–246, 250, 254–256, 259
 Relationalism, 3
 Relational ontology (or relational view), 127, 187, 236, 259
 Relational properties, 236, 238, 240, 248, 254, 255, 258–260
 Relational-transformational (systemic) emergence, 6, 236, 237, 246
 Resolutive-compositional method, 192, 195–197
 Resultant Hamiltonian, 166
 Role functionalism, 15, 16
 Rules of philosophizing, 197, 198

S

Scale, 4, 41–44, 48, 66, 140, 141, 144, 145, 186, 187, 221, 224, 249, 257, 259
 Scale free network, 58
 Scaling relationships, 38, 40, 45, 53–57
 Schrödinger equation, 154, 155, 162, 168, 201, 206
 Second-order relations, 244
 Self maintenance, 5, 86, 93, 104, 226
 Self-organization, 215, 225–226, 246, 251
 Stability, 69, 71, 111, 116–123, 125, 127, 128, 130, 132, 170, 205, 222, 242
 Stability conditions, 257

- Stability problem, 116
Statistical explanation, 76
Statisticalism (or statisticalist), 76
Stereochemistry, 148
Stochastic mechanism, 68, 200
Strong emergence, 231, 232
Structural view of chemical bonding, 174
Superposition, 166, 198
Supramolecular organisation, 153
Symmetry, 155–157, 165, 203, 204, 257
Synchronic emergence, 6, 218, 230
- T**
- Temporal synchrony, 13, 15, 18
Tendency, 45, 56, 75, 140, 148
Theory-reduction model, 2, 255, 256
Thermodynamic hypothesis, 112, 114, 120, 131
Thermodynamic openness, 92
Thermodynamic properties, 116, 143, 144
Thermodynamics, 2, 3, 5, 49, 51, 87, 89, 93, 110–135, 143, 203, 205
Thick reaction mechanism, 145, 147
Thin reaction mechanism, 145, 147
Top-down intervention, 19
Topological explanations, 74
Traditional mechanistic philosophy (core six features), 3
- Transformation, 40, 67, 75, 113, 145–147, 158, 187, 222, 227–231, 236, 237, 239, 240, 243, 244, 246, 249, 251, 252, 260
Transformative relations, 187, 259
Transient state, 115, 118, 123–127
- U**
- Ultimate cause(s), 62, 67, 76
Universal laws, 3, 56, 180, 188
- V**
- Valence bond, 174, 175
Variational tendencies, 75
Vitalism, 250, 251
- W**
- Wavefunction, 154, 156, 157, 162, 166
Weak emergence, 231, 232
Whole, 16, 23, 24, 31, 32, 34–37, 39, 44, 45, 48, 52, 62, 65, 67, 72, 86–88, 92, 94, 112, 114, 125, 152, 181, 192, 194, 202–204, 214, 223, 231, 235–237, 239–246, 249, 250, 253, 255, 256, 259, 260
Work (to perform), 86, 89, 90, 131, 134