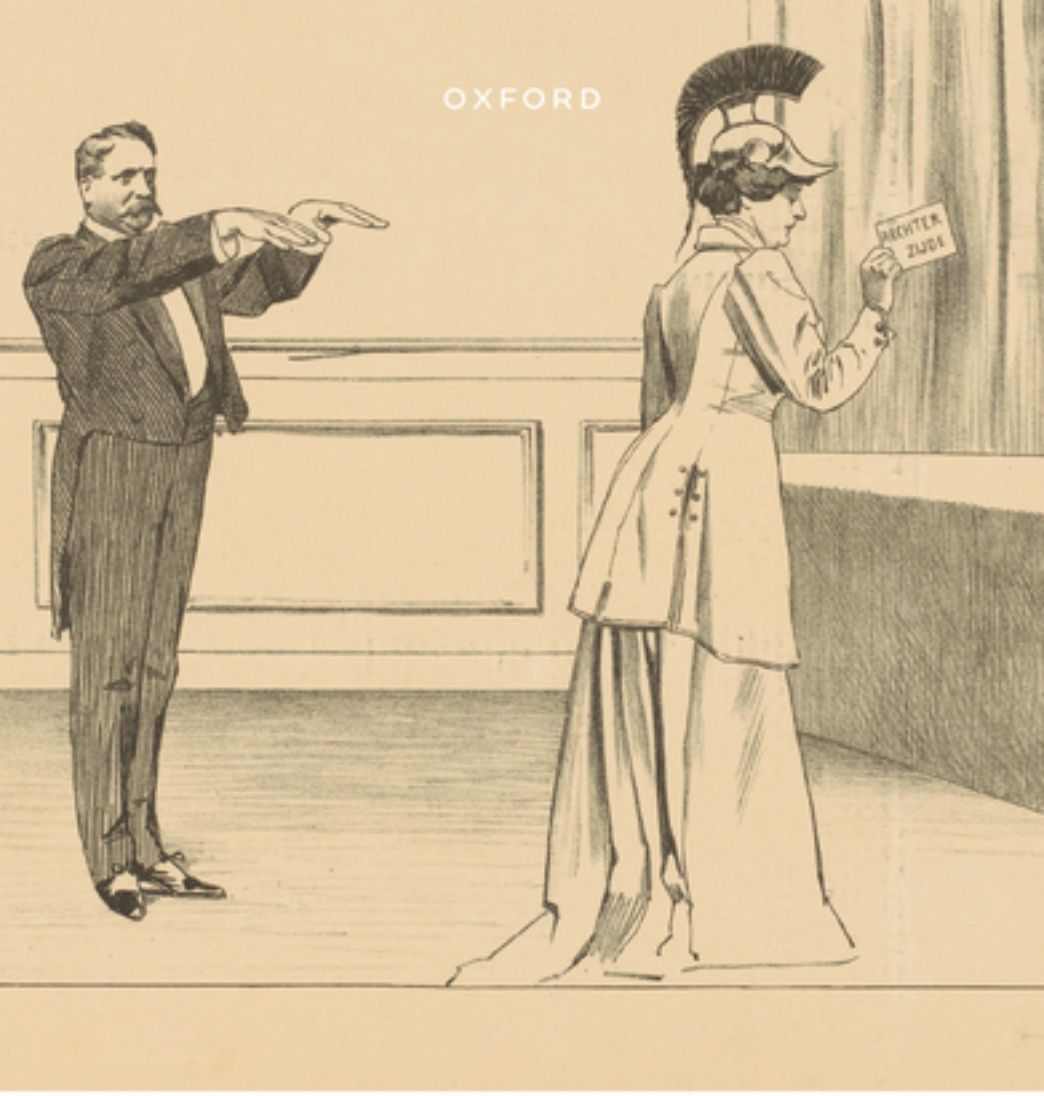

PROTECTING MINDS

The Right Against Mental Interference

THOMAS DOUGLAS

NEW TOPICS IN APPLIED PHILOSOPHY

Protecting Minds

NEW TOPICS IN APPLIED PHILOSOPHY

Series editor: Kasper Lippert-Rasmussen

This series presents works of original research on practical issues that are not yet well covered by philosophy. The aim is not only to present work that meets high philosophical standards while being informed by a good understanding of relevant empirical matters, but also to open up new areas for philosophical exploration. The series will demonstrate the value and interest of practical issues for philosophy and vice versa.

ALSO PUBLISHED IN THE SERIES

Manipulation
Its Nature, Mechanisms, and Moral Status
Robert Noggle

Independence for Children
Matthew Clayton

What's Wrong with Lookism?
Personal Appearance, Discrimination, and Disadvantage
Andrew Mason

Disability Through the Lens of Justice
Jessica Begon

Exploitation as Domination
What Makes Capitalism Unjust
Nicholas Vrousalis

Protecting Minds

The Right Against Mental Interference

THOMAS DOUGLAS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.

British Library Cataloguing in Publication Data

Data available

Library of Congress Control Number: 2025948177

ISBN 9780192883254

DOI: 10.1093/9780191979651.001.0001

Printed and bound by
CPI Group (UK) Ltd., Croydon, CR0 4YY

The manufacturer's authorised representative in the EU for product safety is
Oxford University Press España S.A. of Parque Empresarial San Fernando de Henares,
Avenida de Castilla, 2 – 28830 Madrid (www.oup.es/en or product.safety@oup.com).
OUP España S.A. also acts as importer into Spain of products made by the manufacturer.

For Jocelyn Grace Douglas, 1949–2025

Acknowledgements

I would like to thank the many people who have helped me with my work on this book. Sasha Arridge, Gabriel De Marco, Jonas Hertel, and Martin Wilkinson provided detailed, challenging, and insightful written comments on drafts of the whole book, in each case leading to significant improvements. In addition, Justin Bernstein, Joe Bowen, Lisa Forsberg, Moti Gorin, Viktor Ivanković, Max Kiener, Robert Noggle, Jesper Ryberg, Peter Schaber, and Henry Schiller gave me very helpful written or oral comments on (in some cases multiple) individual chapters.

I also benefitted from feedback from audiences at numerous events, namely: 'Rethinking Bioethics for the 21st Century' (Oxford, May 2019); 'Technology and Manipulation' (King's College London, July 2019, organised by Massimo Renzo); 'Mental Autonomy in the Digital Age' (Chatham House [online], December 2020, organised by Kate Jones, Harriet Moynihan, and Jacqueline Rowe); 'The Ethics of Influence' (Zurich, September 2022, organised by Peter Schaber and Holger Baumann); 'The Zurich Global Health Forum' (Zurich [online], April 2023, organised by Nikola Biller-Andorno and Tania Manríquez Roa); 'The Ethics of Persuasive Digital Technologies' (Oxford, May 2023, organised by Max Kiener); 'Rights for the Mind' (Amsterdam, September 2023, organised by Gerben Meynen and Vera Tesink), 'The Zagreb Applied Ethics Conference' (Zagreb, September 2023, organised by Viktor Ivanković); and 'The Ethics of Influence' (TU Delft, July 2024, organised by Michael Klenk). I received further valuable feedback at departmental talks at Liverpool Hope University (May 2022); University of Canterbury (February 2023); University of Otago (March 2023); and University of Zurich (May 2023), and I benefitted from discussions both of my own drafts and of the related work of other participants in the Oxford-Zurich Work in Progress Meeting Series on the Ethics of Behavioural Influence and Prediction. I did not imagine when I started work on this book that I would end my work on it surrounded by a lively community of scholars interested in similar topics, but this latter meeting series has provided me with just such a community, and I'm very grateful for it.

Several people helped me through one-on-one discussions of particular aspects of the book; I recall particularly helpful chats with Gabriel De Marco,

Lisa Forsberg, Faustino García de la Torre García, Nora Hertz, Marcello Ienca, Alex Kaiserman, Michael Klenk, Muriel Leuenberger, Sjors Ligthart, Allan McCay, Gerben Meynen, Andrew Moore, Robert Noggle, Massimo Renzo, Peter Schaber, Michael-John Turp, and Martin Wilkinson, and there were no doubt others that I am forgetting. I am especially indebted to Gabriel De Marco, from whom I have learned a huge amount through multiple long discussions in the course of my work on this book, and to Robert Noggle, whose outstanding recently published book on manipulation served as a model for me to (unfortunately optimistically) seek to emulate.

I thank also Sasha Arridge and Mia Richmond for help with proofreading; Sasha Arridge for assistance with referencing; Tess Johnston, Faustino García de la Torre García, Elena Radulescu, and Mia Richmond for help in identifying relevant literature; Louise Chapman for completing the index; Kasper Lippert-Rasmussen, both for advice that helped to secure the publication of this manuscript and for many stimulating discussions on topics related to the subject of this book over the years; and Peter Momtchiloff, April Peake, Aimee Wright and Loga Veera Puviyarasi for deftly guiding me through the review and publication process.

This book project was made possible by a Consolidator Award (ProtMind, grant number 819757) from the European Research Council and ongoing research funding from the Uehiro Foundation on Ethics and Education. It was also stimulated by work supported by an earlier award from the Wellcome Trust (Investigator Award 100705/Z/12/Z). I thank all three funders for their extremely generous support, which has provided me with an unusual level of research freedom. I am especially grateful to Dr Noboru Maruyama at the Uehiro Foundation for the strong and unwavering support that he has shown me since I joined the Uehiro Centre for Practical Ethics as a doctoral student almost exactly twenty years ago.

The biggest contribution to my work on this book has been made by my wife, Katrien. Katrien made the space for me to focus on my writing amidst a very busy phase of family life, often sacrificing her own work and interests in order to do so. She also distracted me when I became too obsessed; made me smile when I became fed up; and gave me an immense amount of love, support, and joy. Our children Eleonora and Edmund have also been a constant source of entertainment, wonder, and companionship throughout my work on this book, and I thank them for their love and for never being more than mildly annoyed when my research took my attention away from them.

This being my first book, I have not previously had the opportunity to thank my parents in public writing, and I'm very glad to have the opportunity

to do so. Their encouragement of my curiosity, argumentativeness, and esoteric interests played a crucial role in fostering my love for philosophy. They also passed on a determination and self-confidence that allowed me to pursue that love, despite my having started out on a very different career trajectory.

I dedicate this book to my mother, Jocelyn Douglas, who died shortly after I submitted the full manuscript. Mum, you were an inspirational woman of action, courage and kindness; you loved us fiercely and gave yourself to us without restraint; and you've been my biggest source of strength. Losing you over the months while I was finalising this book felt like losing a large part of myself, and I miss you so much.

Contents

1

Motivating the Project

The Case for Examining the Right Against Mental Interference*

If we possess any moral rights at all, the moral right against interference with our bodies is surely among them. Though there is little agreement about what to call the right—the right 'to bodily integrity' and 'against bodily trespass' are perhaps the most common monikers—its existence is widely endorsed, and indeed is often simply taken for granted.[1] That we possess this right, many would hold, helps to explain why it is generally morally impermissible to conduct medical procedures on people without their consent, to physically assault them, or to subject them to unwanted touching.

Do we also possess an analogous moral right against interference with our *minds*—a right against *mental interference*? It might seem that we must, for our minds are normally at least as important to us as our bodies, suggesting that they deserve similar moral protection. Moreover, some considerations that have been thought to ground a right against bodily interference seem also to be capable of grounding an analogous right against mental interference. For instance, the right against bodily interference is often thought to derive from our self-ownership (Thomson 1990) or personal sovereignty (Archard 2008; Ripstein 2006), but ourselves and our persons surely include our minds as well as our bodies, so it may seem that if these considerations ground a right against bodily interference, they could also ground a mental analogue of that right.[2]

* Several of the points made in this chapter are made also in Douglas and Forsberg (2021), which is cited below where the points are significant. I thank Joseph Bowen for some helpful suggestions regarding the discussion of rights in this chapter.

[1] For defence of the view that the right is often taken for granted, see Douglas and Forsberg (2021). For examples of the 'trespass' terminology, see Thomson (1990) and Archard (2008). For discussion of the various rights that have been referred to as rights to bodily integrity, see Viens (2020) and Dekkers et al. (2005). Note that the right against bodily interference is sometimes regarded as part of a broader right to bodily (or more generally personal) autonomy or sovereignty. See, for example, Feinberg (1989).

[2] Jan Christoph Bublitz and Reinhard Merkel (2014, p. 62) make a similar point: 'In the wake of Locke, libertarians believe that persons have property rights in their body; persons literally

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0001

The moral right against interference with our minds has, however, not received the same level of attention or acceptance as its bodily analogue. That is not to say that it has received *no* attention. As we will see later in this chapter, legal institutions and legal philosophers have advocated the recognition of something close to a *legal* right against mental interference, and moral philosophers are also beginning to entertain the possibility that we might enjoy a parallel moral right.[3] Moreover, many moral philosophers have defended broader rights that plausibly ground or contain something like a right against mental interference. The right to freedom of thought is the clearest example,[4] but others may include the right to personal autonomy, rights of self-ownership, the right to respect, and the right against paternalistic treatment.

There is, then, a growing body of philosophical work that either purports to support, or could plausibly be marshalled to support, the existence of a moral right against mental interference. Nevertheless, whether

own (the physical part of) themselves. Ownership discussions focus on the relation of persons to their bodies, their liberty (e.g. vis-a-vis slavery) and the fruits of their labor. But what is even more constitutive of a subject than her body is her mind. So, whoever grants self-ownership of persons over their bodies has a compelling reason to concede self- ownership over minds.'

[3] Authors who have defended or entertained the possibility that we might possess a moral right against mental interference, or something close to it, include Neil Levy (2007, pp. 179–80), who mentions a possible liberty right over the mind; Jared Craig (2016, p. 112), who defends a yet-to-be-fleshed-out right to mental integrity, which would protect 'autonomous human agency'; Adina Roskies (2016), who considers a possible right to cognitive liberty, which she suggests would protect against manipulation and thought-control; Jesper Ryberg (2019, Ch. 3), who considers various possible formulations of a right to mental non-interference; Frederick Schauer (2020, p. 77), who, in a generally sceptical discussion of the right to freedom of thought, concedes that we may possess a right against 'influences on our thinking [that] are unknown to us, or in some other way circumvent our presupposed rational capacities'; Elizabeth Shaw (2022, p. 1418), who asserts a moral right to mental integrity that 'can be infringed by intentionally interfering with a person's mental states through nonrational means'; Andrea Lavazza and Rodolfo Giorgi (2023, p. 2), who defend a right to mental integrity, where mental integrity is understood as including 'the ability to instantiate one's mental/brain states . . . without interference'; and Will Ratoff (2024), who elucidates a right to mental autonomy, understood as a right against nonrational interference with the right-holder's thinking. In previous work, Lisa Forsberg and I have also suggested (without assessing) arguments for a right to mental integrity, understood as a right that protects against significant, nonconsensual interferences with the mind (Douglas 2014; Douglas & Forsberg 2021), and Bartlomiej Chomanski (2023) has recently endorsed these arguments and this right, and argued that one can derive, from this right, a right against certain types of distraction.

[4] J. S. Mill claimed that we possess a right to 'liberty of conscience, in the most comprehensive sense; liberty of thought and feeling; absolute freedom of opinion and sentiment on all subjects, practical or speculative, scientific, moral, or theological' (1909, p. 23). I will return to the moral right to freedom of thought, and its relationship to the putative right against mental interference, in the next chapter.

there is such a right, and, if so, precisely what protection it provides to our minds, remains very much an open question. So, too, do many related questions. What, exactly, is mental interference, and how is it to be distinguished from innocuous forms of mental influence, such as presenting people with sincere and sound arguments for changing their mind? Do all mental interferences infringe the right against it, or only some subset of these interferences? And if the latter, which subset, exactly? Would a right against mental interference provide any protection that is not already provided by the right against *bodily* interference? And what theoretical basis can be given for the right?

By confronting these questions among others, I aim, in this book, to investigate and elucidate the right against mental interference and the arguments for and against it. I will argue that we *do* possess such a right. And I will draw out the implications of the right for a range of different types of mental influence, ranging from the compulsory administration of mind-altering drugs and the nonconsensual deployment of brainwashing techniques, to more familiar and mundane influences that would commonly be regarded as marketing techniques or nudges.

In this chapter, I set the scene for the argument to come by introducing and motivating the thought that we possess a moral right against mental interference and by explaining why I think this putative right warrants greater attention than it has so far received.

The chapter is organised as follows. In Section I, I introduce and motivate the view that we all possess a moral right against bodily interference. I then, in Section II, motivate the thought that we also possess an analogous moral right against mental interference. The plausibility of the thought that we possess this right already gives us good reason to further elucidate the right and the arguments for and against it. In Sections III and IV, I highlight two further reasons to do so. The first reason is that the mind is highly vulnerable to interference, suggesting that, if we do possess a right against mental interference, it might be of considerable practical importance, since it might rather frequently be infringed. The second reason is that there have been recent moves to recognise a *legal* right against mental interference, or something close to it, and whether these moves are advisable may well depend on whether we enjoy a corresponding moral right.

By the end of Section IV, I hope to have established that the putative moral right against mental interference deserves greater attention than it has so far received. In Section V, I give a brief plan of the attention I will give it in the remainder of the book.

I. The Right Against Bodily Interference

Suppose that an intruder creeps into your bedroom while you are sleeping, pierces your skin with the needle of a syringe, and, without your noticing, injects the contents of the syringe into your triceps. And suppose that you knew nothing in advance of his plan to do this.[5]

Clearly the intruder has wronged you.[6] As a result, you have a legitimate moral complaint against him. But *how* exactly has he wronged you? What is the nature of your legitimate complaint against him? This is less clear. Perhaps the substance that he has injected will cause you to feel queasy, or light-headed or weak, and perhaps he has thereby wronged you by causing you to experience an unpleasant state. But suppose that all he injected was a tiny amount of sterilised saline, and that neither the saline nor the process of injecting it has any noticeable effect on you. Still, the intruder seems to have wronged you. How?

Perhaps he has wronged you merely by entering your bedroom without your permission. Perhaps this alone amounts to a trespass on your property or an invasion of your privacy. But this cannot be the whole story, for surely the wrong the intruder perpetrates against you is a greater wrong than the wrong he would have committed had he entered your bedroom, but without injecting you with anything. His injecting you seems to have made a moral difference.

One plausible explanation of the difference made by his injecting you would invoke the concept of bodily interference. The intruder wrongs you by interfering with your body without your consent. The intruder—like everyone else—falls under a standing moral duty, owed to you, not to interfere with your body. You may be able to release others from aspects of this duty, for example, by consenting to a particular interference. Thus, by consenting to the removal of your appendix, you may release your surgeon from the duty not to interfere with your body in this particular way. The surgeon can then perform the appendicectomy without wronging you. But in the case of the intruder, the duty remains fully in force; you have not released the intruder from any aspect of it. So when the intruder injects you with saline, he acts contrary to the duty and thereby wrongs you.

[5] This case also appears in Douglas and Forsberg (2021).

[6] As I explain below, I take it that you are wronged, or treated wrongfully, whenever someone fails to live up to a *pro tanto* (i.e. defeasible) duty, owed to you. Because the duty is only a *pro tanto* one, wrongful treatment is not always impermissible or (as I will take to be equivalent) wrong, all things considered.

Here's another, perhaps more familiar, way of saying what arguably amounts to the same thing, this time framed in terms of rights rather than duties. You have a moral right that others, including the intruder, not interfere with your body. You may be able to partially waive this right, for example, by consenting to an interference. This may prevent the consented-to intervention from infringing the right. But in the case of the intruder, you have not waived any aspect of your right, so the intruder infringes it, and thereby wrongs you.

I suggested that the latter, rights-based explanation is equivalent to the former, duty-based explanation, but in fact, that would be somewhat controversial. On many accounts of rights, the intruder infringes a right of yours only if he fails to fulfil some *special* kind of duty that he owes to you. Perhaps the duty has to be (near) absolute,[7] a trump,[8] or at least enforceable by the state.[9] Yet I did not, in my duty-based explanation, characterise the duty in any of these ways. I nevertheless claim that the duty-based explanation and the rights-based explanation are equivalent. This is because I subscribe to a weak account of rights on which having a right entails nothing more than that something is owed to you, by which I mean that someone else falls under a *pro tanto* duty—a duty that has some, though not necessarily decisive, normative force—and that duty is directed towards you.[10] This duty need not be (near) absolute, a trump, enforceable, or in any other way special. It is, I think, somewhat plausible that the intruder infringes a right of yours even on some stronger accounts of rights. For example, it seems to me plausible that the intruder acts contrary to an *enforceable* duty, owed to you, not to interfere with your body. But in what follows, when I claim that someone infringes your moral right against bodily interference, I mean to imply only that she fails to fulfil some *pro tanto* moral duty, owed to you, not to interfere with your body.[11] Because, on my view, having a right

[7] Robert Nozick (1974) is often understood as holding that the duties corresponding to rights are absolute (see, e.g. Francis & Francis 1976, p. 634). Others attribute to Nozick the slightly weaker view that duties corresponding to rights are *almost* absolute, on the basis that, in a footnote (Nozick 1974, p. 30), Nozick leaves open the possibility that rights might be permissibly infringed in order to avoid 'catastrophic moral horror' (see, e.g. Arneson 2011; Vallentyne 2011).

[8] See, for example, Dworkin (2013, p. 6). I take the claim that a duty is a 'trump' to mean that it can only be defeated by other duties that are also trumps.

[9] See, for example, James (2003) and Flanigan (2019).

[10] I endorse also the reverse entailment: being the object of a directed *pro tanto* duty entails possessing a right.

[11] The 'owed to you' is doing some work here. It may be that we can have duties owed to no-one, or no-one in particular. For example, I may have a duty, owed to no-one, not to destroy a beautiful tree on a small island that will never be inhabited, or even seen, by any sentient creature. Duties that are owed to particular beings plausibly have implications that such impersonal

entails only a *pro tanto* duty, not an absolute one, my view allows that it may sometimes be permissible to infringe the right against bodily interference, say, in order to prevent the right-holder or a third party from suffering serious harm. For example, it might be permissible to push a person out of the way of a fast-approaching tram, even if doing so would (as I believe) infringe their right against bodily interference. Indeed, I happen to think not only that it *could* be permissible to infringe a person's right against bodily interference, but that it sometimes *is* permissible to do so. However, I do think that infringing this right always *wrongs* the right-holder, for I take wrongings, like rights infringements, to consist in the breach of a directed *pro tanto* duty.

II. The Right Against Mental Interference

Though it is rarely discussed in detail or fully specified, the existence of a moral right against bodily interference is widely accepted in moral, legal, and political philosophy (Douglas & Forsberg 2021). It has played an especially prominent role in discussions of the wrongfulness of nonconsensual sex, organ harvesting, and medical intervention.[12] By contrast, the moral right against *mental* interference, and the corresponding *pro tanto* moral duty not to interfere with others' minds, is much less widely endorsed in scholarly work.[13] Some authors have, however, begun to wonder whether we enjoy something like this right or bear something like this duty.[14] And I think they have been correct to do so.

For one thing, as I have already noted, some standard attempts to ground the moral right against bodily interference—those appealing to self-ownership and personal sovereignty—seem capable of grounding a moral

duties do not have. For example, if I fail to live up to a duty owed to you, I will plausibly owe you compensation, whereas if I fail to live up to an impersonal duty, it is less clear that compensation could be a fitting response, in part because it is not clear to whom any compensation should be given. Similarly, if I fail to live up to a duty owed to you, it is plausible that you have *special* grounds for complaint against me, whereas if I fail to live up to an impersonal duty, it is doubtful that anyone has a special ground for complaint.

[12] T. M. Wilkinson (2011, p. 16) notes that '[t]he right to bodily integrity ... is almost entirely uncontroversial and often considered of great weight'.

[13] Again, I do not assume that this duty must be absolute, a trump, enforceable, or in any other way special.

[14] As we will see in the next section, there has also been some discussion of a *legal* right against mental interference, or something close to it.

right against mental interference too, given the central role played by the mind in both the person and the self.[15]

For another, the view that we possess a moral right against mental interference is able to explain some widely held intuitions (Douglas & Forsberg 2021). Consider a variant of the above mentioned intruder case. An intruder creeps into your bedroom while you are sleeping and injects you with a substance. But this time, the substance is not biologically inert. Rather, it is a stimulant drug similar to methylphenidate, the active ingredient in Ritalin. Following injection, it enters the brain and stimulates certain brain regions in a way that tends to enhance a person's ability to focus. Suppose that the intruder injects you with the stimulant for the purpose of enhancing your attention. And suppose that he is successful in realising this goal: for the next two days, you are more focussed on your work than would otherwise have been the case.

As in the initial version of the case, it is plausible that the intruder has infringed your moral right against bodily interference. Indeed, I think the intruder has interfered with your body in at least three ways: by piercing your skin with the needle, by injecting a drug, and by chemically altering those neural states that mediate your attention.[16] Moreover, and crucially, you have not waived your right against bodily interference in relation to any of these interferences by, for example, consenting to them in advance.

However, it seems doubtful that bodily interference is the *only* problematic feature of the intruder's intervention in this case. The intervention is also problematic, it might seem, by virtue of its impact on your mind. For suppose the intruder had injected you with a different drug—one that was similarly biologically active, in the sense that it produced physiological and biochemical changes similar in magnitude to those produced by the stimulant, but which did not at all affect the brain or nervous system or, therefore, the mind. Suppose that the impact of this drug was rather confined to some other organ—the liver, say—and that neither the drug itself, nor the process via which it is introduced into your body, had any effect on your mental life;

[15] It has, indeed, been argued that one of the most influential accounts of self-ownership—John Locke's—*only* directly supports rights over the mind, not rights over the body. As Jeremy Waldron (1990) notes, Locke asserts property rights in the *person* and operates a technical, psychological account of personhood on which the body is merely a home for the person, not a part of it.

[16] I will return to the question of what, exactly, counts as bodily interference in the next chapter.

your mental life is exactly as it would have been had the intruder never injected you with this substance.

It seems to me that injecting the liver-altering drug is in at least one non-trivial way less wrongful, or at least less intrusive, than injecting the stimulant drug. And one way—though admittedly not the only way[17]—to account for this difference would be to posit a moral right against mental interference. This right would plausibly be infringed by the brain-stimulating intruder, who seemingly interferes with both your mind and your body. However, it would not be infringed by the liver-altering intruder, who interferes only with your body.

III. The Vulnerability of the Mind

I have been suggesting that there are both theoretical and intuitive reasons to take seriously the possibility that we each possess a moral right against mental interference. Why, then, has the right been relatively neglected?

Various explanations are available.[18] Perhaps it has been thought that a right against mental interference would be toothless, since, given the unobservability of others' minds, infringements of it would almost always be difficult to identify with any confidence.[19] Or perhaps it has been thought that, since we are constantly influencing, and arguably interfering with, one another's minds, recognising a right against mental interference would implausibly imply that we are constantly infringing one another's rights.[20]

Whether or not these considerations help to explain why the right against mental interference has received less attention than its bodily analogue, I do not think that either *justifies* this relative neglect. After all, it would be widely accepted that there is a duty, or at least a reason, not to cause others pain, but identifying pain also requires us to observe the contents of others' minds. And it is not obvious that recognising a right against mental interference would result in an unacceptable proliferation of rights infringements. It could be argued that many of the most common ways in which we affect others' minds count as mere mental *influences*, not as mental *interferences*.

[17] I will consider alternative ways of accounting for what is problematic about this interference in Chapter 4.

[18] For discussion, see Douglas and Forsberg (2021).

[19] For discussion of the analogous thought in relation to the *legal* protection of the mind, see Bublitz and Merkel (2014, p. 52).

[20] I will consider this worry in detail in Chapter 8.

It could also be argued that only a subset of mental interferences infringe the right against mental interference. And finally, it could be argued that we should simply accept that we *are* infringing one another's rights much more often than we generally think.

Another possible explanation of—and perhaps justification for—the relative neglect of the right against mental interference appeals to humanity's historical confidence that the mind is insusceptible to interference, or at least, to the sort of thoroughgoing interference that would seem most morally troubling (Blitz 2010, pp. 1051–52; Bublitz & Merkel 2014; Douglas & Forsberg 2021; Ienca & Andorno 2017; Schauer 2020, p. 73). In discussing the need for more effective legal protection of the mind, Jan Christoph Bublitz and Reinhard Merkel suggest that such protection has not previously been forthcoming because 'traditionally, the mind has not been conceived as an entity vulnerable to external intrusions and hence in need of legal protection' (Bublitz & Merkel 2014, p. 61); instead, there has been a firm belief in the 'untouchable absoluteness of freedom of the mind' and 'the factual invincibility of the mental realm' (Bublitz & Merkel 2014, p. 65).[21] Perhaps the mind has, for the same reasons, not been considered in need of *moral* protection.

However, any belief in the invincibility of the mind is at odds with the facts. Humans have long been vulnerable to rhetorical techniques that can—as Plato (2019) famously argued—be used to mislead and exert power over the listener. They have also been vulnerable to forms of outright mental trickery that are capable of exerting significant, if short-lived, effects on a person's mental states: consider, for example, the use of distraction by salespeople, advertisers, con artists, and magicians.[22] And they have been vulnerable to more extensive campaigns of political and religious indoctrination in which one or a few individuals appear to obtain an unusual level of control over large swathes of their victims' mental lives. The nineteenth and twentieth centuries witnessed the development of some even more extreme forms of mental influence, including psychopharmacological treatments

[21] Bublitz and Merkel (2014, p. 64) also cite evidence that delegates involved in drafting the Universal Declaration on Human Rights subscribed to this view; for example, the Belgian delegate is reported to have held that 'it would be unnecessary to proclaim that freedom [of conscience] if it were never to be given an outward expression; if it were intended, so to speak, only for the use of the inner man' [UN General Assembly (1948, p. 395); cited in Hammer (2018, p. 34)]. Elsewhere, Ienca and Andorno (2017, p. 1) entertain similar thoughts: 'While the body can easily be subject to domination and control by others, our mind, along with our thoughts, beliefs and convictions, [have until recently been] to a large extent beyond external constraint.'

[22] For discussion of some of the most striking applications of distraction, see Simons and Chabris (2010).

for psychosis and mood disorders, psychosurgical interventions such as the frontal leucotomy, and a range of psychological techniques often referred to collectively as 'brainwashing'. And now, in the early decades of the twenty-first century, we are witnessing the rapid development of subtler forms of mental influence—and arguably interference—to which I will return shortly. Thus, whatever role historical belief in the invulnerability of the mind may play in explaining neglect for the right against mental interference, no one could credibly invoke such invulnerability to justify this neglect today.[23]

Indeed, it seems to me that the substantial and increasing vulnerability of the mind to outside interference makes the assessment of the case for a moral right against mental interference a matter of considerable practical importance; it is possible that our right against mental interference is very often infringed, and that it may be even more frequently infringed in the future.

The existing interventions that most plausibly infringe the right against mental interference are what we might call *nonconsensual neurointerventions*: interventions that influence the mind through the physical (including chemical) modulation of brain states, and that are performed without the valid consent of those subjected to them.

Nonconsensual neurointerventions are a feature of psychiatric practice in many jurisdictions. Mental health legislation standardly allows for brain-active pharmaceuticals to be imposed on psychiatric patients deemed to pose a risk to themselves or others, and in some jurisdictions, these treatments can be imposed even in cases where those patients retain decision-making capacity.[24] Nonconsensual neurointerventions are also, though less commonly, employed in the context of criminal justice. For instance, some US states and European countries allow for the imposition of anti-libidinal hormonal agents—often referred to as 'chemical castration'—on some sex offenders, either as part of a criminal sentence or as a condition of parole or early release (Forsberg & Douglas 2016; Thibaut et al. 2010). In criminal justice contexts, it might be held that the individual subjected to the intervention has rendered herself liable to it; by committing a serious crime, the individual may have forfeited the rights that would otherwise have protected him against it. However, in psychiatric contexts, it is very doubtful that any rights have been forfeited.

[23] This is a point also made by Blitz (2010), Bublitz and Merkel (2014), and Ienca and Andorno (2017).

[24] This is the case, for instance, in England and Wales. See, for example, Zigmond (2017).

Nonconsensual neurointerventions are perhaps the clearest and starkest examples of interventions that would infringe the right against mental interference. But such a right might also be infringed by many *psychological* forms of mental influence. Consider, for example, aversion therapy, subliminal imagery, hypnosis, and the other extreme forms of mind control commonly referred to as 'brainwashing'. Consider also the less extreme but far more prevalent forms of online influence often referred to as 'persuasive digital technologies'. Though this class of interventions has no accepted, precise definition, it is often said to include measures such as Amazon's one-click-to-buy feature (which facilitates impulse purchases), clickbait headlines (designed to create an urge to follow a link), bottomless newsfeeds (intended to prolong engagement by exploiting a fear of missing out), and randomised rewards, for example, in the form of 'loot boxes' in computer games (intended to prolong game play or increase player spending). There is growing disquiet regarding persuasive digital technologies, even within the tech community, with the predominant concern being that they are used to distract individuals from pursuing their autonomously chosen, authentic, or worthwhile long-term goals (Bhargava & Velasquez 2021, pp. 328–33; Harris 2016; Solon 2017; Susser et al. 2019, esp. p. 9; Tran 2016; Wong 2017). Yet even when used to advance such goals, it is reasonable to worry that persuasive digital technologies interfere with our mental lives and thereby infringe a right against mental interference.[25]

Finally, consider what are sometimes called heuristic-triggering nudges. We often decide how to act not through slow, careful, and conscious deliberation, but by employing simple, quick, and sometimes subconscious heuristics such as 'choose what is most salient', 'stick with the default', or 'listen to people you recognise'. Heuristic-triggering nudges (henceforth just 'nudges') seek to influence our decisions or choices by exploiting such heuristics.[26] They have, since receiving their canonical defence in Thaler and Sunstein's (2008) book *Nudge*, been widely deployed by policymakers. Examples

[25] A similar point is made by Daniel Susser et al. (2019, p. 11, emphasis in original): 'it is harmful to manipulate someone even in an effort to lead them more effectively toward *their own self-chosen ends*. That is because the fundamental harm of manipulation is to the process of decision-making, not its outcome'.

[26] Not all interventions commonly referred to as 'nudges' trigger heuristics. Bart Engelen (2019, p. 219) distinguishes 'heuristics-triggering nudges' from 'heuristics-blocking nudges', which would include 'reminders, prompts and cool-down periods', and which serve to prevent deliberative reasoning from being 'overrun' by heuristics. Presenting simplified information to reduce the burdensomeness of deliberative decision-making would be another example of a heuristics-blocking nudge (see, for discussion, Münscher et al. 2016, p. 515).

include the placement of healthy foods in prominent locations in cafeterias, the serving of meals on smaller plates, the removal of distinctive labels and logos from cigarette packages, and the painting of eyes above bicycle racks (to discourage theft).

Nudges have become the target of much philosophical discussion and numerous ethical objections. To my knowledge, none of these objections adverts explicitly to the phenomenon of mental interference, or to the right against it. But nudges are intended to alter, and, when successful, do in fact alter, mental states or events—typically decisions, choices, or intentions. It may not be too much of a stretch to regard them as forms of mental interference. And this at least raises the possibility that they might infringe a moral right against mental interference.

I will return to consider neurointerventions, persuasive digital technologies, and nudges in more detail in Chapter 7, where I will argue that some such interventions indeed infringe the right against mental interference. Here, my purpose is merely to highlight that there exist forms of influence—in some cases prevalent forms of influence—that *plausibly* infringe the moral right against mental interference. This is enough to establish that the right has a strong practical claim to philosophical attention.

IV. Legal Significance

The moral right against mental interference also has another claim to philosophical attention: such attention is needed to inform our assessment of some recent legal developments.

Just as the moral right against mental interference has received less attention than its bodily analogue, so too, the *legal* right against mental interference has received less attention than *its* bodily analogue (Douglas & Forsberg 2021). The law in many jurisdictions clearly recognises a right against bodily interference.[27] English courts have, for example, held that 'every person's body is inviolate',[28] and that '[e]very human being's right to life carries with it, as an intrinsic part of it, rights of bodily integrity and autonomy'.[29] Similarly,

[27] For a discussion of the right to bodily integrity in law, see Herring and Wall (2017). For a classic characterisation of the right as a right against bodily interference, see Feldman (2002, p. 241).

[28] *Collins v Wilcock*, [1984] 1 W.L.R. 1172 (1984).

[29] L. J. Walker in Re A (conjoined twins), [2001] 2 WLR 480 (2000). See also *R (on the application of Justin West) v The Parole Board*, [2002] EWCA Civ (2002).

the New York Court of Appeal famously held that 'every human being of adult years and sound mind has a right to determine what shall be done with his own body',[30] while the Supreme Court of Canada has asserted that '[e]very patient's right to bodily integrity encompasses the right to determine what medical procedures will be accepted'.[31]

Indeed, the legal right against bodily interference is so well established that it is a staple of undergraduate textbooks (see, for a summary, Douglas & Forsberg 2021).

By contrast, the legal right against *mental* interference has only patchy recognition (Douglas & Forsberg 2021). Some legal documents do explicitly recognise a right in the vicinity of a right against mental interference. For example, the European Charter on Human Rights recognises rights to bodily *and mental* integrity in the same breath, regarding both as aspects of a broader right to personal integrity (*Charter of Fundamental Rights of the European Union* 2012). However, the content of this right to mental integrity remains obscure. Moreover, it has achieved nothing comparable to the widespread recognition achieved by the legal right against bodily interference and has played a negligible role in case law.

As noted above, some legal scholars have argued that this ought to change.[32] In their seminal article, 'Crimes Against Minds', Jan Christoph Bublitz and Reinhard Merkel (2014, pp. 58, 60) defend the explicit recognition of a right to mental self-determination which they take to be already implicit in the law, and which, they posit, would include a right to 'freedom from mental manipulations' and 'severe [mental] interferences by the state and third parties'.[33] Similarly, Susie Alegre (2017, p. 224) argues for greater recognition that the right to freedom of thought asserted by Article 9 of the European Convention on Human Rights (ECHR) and Article 18 of the International Covenant on Civil and Political Rights already includes a right 'not to have one's thoughts or opinions manipulated'. Numerous others have made similar claims.[34]

[30] *Cardozo J in Mary E. Schloendorff v Society of New York Hospital*, 105 N.E. 92 (New York Court of Appeals 1914).

[31] *Ciarlariello v Schacter* (1993(2) SCR 119 1993).

[32] For overviews of these calls for change, see, for example, Douglas and Forsberg (2021) and Ligthart, Bublitz et al. (2023).

[33] The right to mental self-determination would also, they think, include a 'positive dimension', which they characterise as a 'freedom *to* self-determine one's inner realm, e.g. the content of one's thoughts, consciousness or any other mental phenomena' (Bublitz & Merkel 2014, p. 60, their italics).

[34] For others arguing that the right to freedom of thought includes, or may include, a right against (something like) interference with thoughts, see Murdoch (2007), who refers

These scholars hold that something like a legal right against mental interference is implied by legal judgments or other accepted legal rights.[35] Their main claims are that this right needs to be made more explicit and better enforced. Others have suggested that the law ought to recognise *new* rights prohibiting certain forms of mental interference. For example, Marcello Ienca and Roberto Andorno, and (separately) Nita Farahany have argued for recognition of a new right to cognitive liberty. Ienca and Andorno (2017, p. 10) take this to include a negative right protecting 'individuals from the coercive and unconsented use' of emerging neurotechnologies,[36] while Farahany (2023, p. 8) takes it to include a 'right to self-determination over our brains and mental experiences', which in turn includes a right against certain forms of manipulation of our minds.[37] Elsewhere, Jasper L. Tran (2016) and Anuj Puri (2021) have each argued for the legal recognition of, respectively, a right to attention and to attentional privacy, in both cases understanding this as including a right against (certain) unwanted intrusions on our attention.

All of these calls could be understood as calls for recognition of a legal right against (certain kinds of) mental interference. However, all leave much work to be done in specifying the content of this right. For example, none offer an account of what, exactly, qualifies as 'mental interference' or—when they prefer this term—'mental manipulation'. Moreover, they all leave it open how, precisely, we ought to *morally* justify the legal rights

to indoctrination; Vermeulen and van Roosmalen (2018), who refer to indoctrination and brainwashing; Bublitz (2014, esp. p. 3), who refers to indoctrination, brainwashing, and thought-manipulation; McCarthy-Jones (2021), who refers to thought-manipulation, thought-interference, and thought-control; and Blitz (2010, pp. 1084–88), who refers to targeting of, intruding on, ruling out, and restricting thought. Gabriel S. Mendlow (2018, p. 2346, also at pp. 2376–83) argues that American case law also already implicitly recognises 'a right of mental integrity, a right to be free from the direct and forcible manipulation of their minds' (emphasis removed from the original). There are difficult and open questions concerning the relationship between mental interference, on the one hand, and mental manipulation on the other hand. But it is plausible to think that there is at least a substantial overlap between these, suggesting that the right against mental manipulation invoked by some of the scholars cited in this note may be at least a close relative of the right against mental interference. I return to this issue in the next chapter.

[35] A point reiterated in Ligthart, Bublitz et al. (2023).

[36] Ienca and Andorno (2017) understand the right to cognitive liberty also to include a positive element—namely, a right to *use* emerging neurotechnologies.

[37] Farahany (2023, p. 165) suggests, for example, that this right would protect against hard-to-resist manipulation: 'If a product becomes actually or nearly impossible to resist, our freedom of action will be hindered and our self-determination and freedom of thought will be put at risk.' Page numbers for Farahany (2023) are for the Kindle version of this book.

whose recognition is being proposed; the proponents of these rights are primarily interested in *legal* arguments for them. One possibility, of course, is that these legal rights might be morally justified by reference to something like a moral right against mental interference.[38] Yet none of these authors takes a clear stance on whether or why this sort of justification might succeed. One reason to determine whether we possess this moral right is, then, that this may help us to morally appraise recent calls for, and indeed moves to recognise, a *legal* right against mental interference, or something close to that right.

V. The Plan

I hope I have convinced you that the moral right against mental interference warrants greater attention than it has, to date, received. One of my aims, in the remainder of this book, will be to begin to give it the attention I think it deserves. My more substantive aims will be to (i) establish the existence of a moral right against mental interference, (ii) specify its scope, and (iii) draw out some of its implications.

My examination and defence of the right against mental interference will proceed as follows. In Chapter 2, I offer a schematic working account of the right against bodily interference, which I then, in Chapter 3, use as a template in developing a working account of the analogous right against mental interference. In Chapter 4, I develop an intuition-based justification of the right against mental interference, and in Chapter 5, I supplement this with a theoretical justification, which seeks to derive the right from our ownership of ourselves.

In Chapter 6, I turn from questions concerning the justification of the right against mental interference to questions concerning its scope. I further specify the scope of the right by refining the preliminary account of the right offered in Chapter 3. This results in a more precise account. However, the account remains abstract; it does not yet provide concrete guidance on which mental influences infringe the right, and which do not. Chapter 7 begins to provide such guidance by drawing out the implications of my account across a range of different cases of mental influence.

[38] It is sometimes held that, in the case of legal *human* rights, legal rights *must* be grounded on moral rights. See, for example, Cruft (2019).

Finally, in Chapter 8, I introduce and reply to two objections to the view that we possess a right against mental interference, and in Chapter 9, I briefly summarise my argument, before raising some of the many important questions about the right against mental interference that I have had to leave unanswered.

2

A Working Account of the Right Against Bodily Interference*

What is the moral right against mental interference, exactly? I conceive of it as a mental analogue of the better-established moral right against *bodily* interference (RABI), and I will approach it via that right. In this chapter, I offer a schematic account of the RABI. In the next chapter, I develop a parallel account of the moral right against mental interference (RAMI). In both cases, I conceive of the right in question as a *moral* right, and, as previously, I understand moral rights to correspond to directed *pro tanto* moral duties.

My accounts of both rights are intended as working accounts—not definitive specifications—and, in the case of the RAMI, my account will be refined over the course of the subsequent chapters. The accounts are also incomplete. I hope they capture many of the core types of cases in which these rights are infringed, but I do not claim that they capture all cases. Thus, my accounts take the form of sets of conditions whose satisfaction I take to be *sufficient* for the infringement of these rights; I do not claim that the conditions are also necessary. There may be ways of infringing the rights against bodily and mental interference that my accounts do not capture.

As well as being tentative and incomplete, my schematic accounts of the rights against bodily and mental interference are intended to be *metaphysically ecumenical*. By this I mean that they are intended to be compatible with a broad range of views regarding the metaphysical relationship between the mind and the physical world, including the brain and body. They are, for example, intended to be compatible with each of physicalism (which I take to be the view that mental states are physical states), emergentism (which I take to be the view that mental states are not physical states, but are grounded in physical states), and dualism (which I take to be the view that mental states are neither physical states nor grounded in physical states). At various points in this book, I will give examples in which one agent alters the

* This chapter draws on some arguments initially published in T. Douglas (2025b).

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0002

mind of another by altering their brain states. These examples presuppose that the mind can be causally affected by the brain and thus presuppose the falsity of some strong forms of dualism. They presuppose, for example, the falsity of Leibnizian mind–body parallelism, according to which all mental events have only mental causes, while bodily events have only bodily causes. However, such examples are consistent with physicalism, emergentism, and some forms of dualism. Moreover, examples aside, much of what I say in these chapters and elsewhere in the book will be compatible even with mind–body parallelism.

There is, however, one important exception to my ecumenism. In addition to the metaphysical debate concerning the relationship of the mind to the physical world, there is debate concerning *which parts* of the physical world serve as the physical realisers of a person's mental states.[1] Most notably, there is the ongoing debate over whether the mind is realised only by the body, or also partly by external cognitive aids, such as our diaries, books, and smartphones.[2] If the latter, then stealing a person's smartphone or rearranging the regularly consulted books on a person's bookshelf might, even in the absence of any neural or other bodily effects, interfere with the mind, perhaps in much the same way as would chemically modulating the person's neural states. For most of this book, I will simply assume that, if the mind is realised by physical entities, those entities are located wholly in the body, so that environmental changes of the sort just described do not, absent further effects, constitute mental interferences. However, I will revisit this assumption in Chapter 8.

I have two further preliminary comments. The first concerns the nature of the beings whose bodies and minds are being interfered with, and whose rights against interference are being infringed. Throughout these two chapters, and most of the rest of the book, I will assume that these beings are adult human persons.[3] I will return briefly in Chapter 9 to consider which other beings might also possess the RAMI, though I will not be able to decisively answer that question.

The final preliminary comment concerns my methodology. Although, as I have said, the RABI is much more widely accepted than the RAMI, it is

[1] There has also been significant discussion of the ethical implications of these metaphysical debates. See, for example, Bublitz and Merkel (2014), Buller (2013), and Glannon (2014).

[2] See, for example, Anderson (2008), Clark and Chalmers (1998), and Levy (2007).

[3] Though I suspect non-humans can also be persons, when I refer to persons in this book, I will always have human persons in mind. This is important because other kinds of persons could have rather different kinds of bodies and minds.

itself under-theorised.[4] Indeed, I know of no fully worked-out account of the right. Thus, in developing my schematic account of the RABI in this chapter, I will be relying heavily on intuition. My aim will be to formalise the RABI that I take to be implicit in our intuitive moral judgements. I will then, in offering an account of the right against *mental* interference in the next chapter, seek to specify that right in such a way that it closely parallels the RABI, while also capturing our intuitions about mental interference.

I refer to 'our' intuitions, but of course I will be relying most heavily on my own intuitions. I hope and predict that many readers will share those intuitions. To those who do not, I can offer the reassurance that the accounts I offer here are intended only as starting points for the normative argument to come. I will go on to defend the view that the right against mental interference, as roughly characterised in the next chapter, is one that we possess. To put it in scientific terms, my main goal in this chapter and the next is to *refine* my hypothesis; most of the *testing* of that hypothesis will come later.

With these qualifications in hand, let me proceed to the task of specifying the RABI. I will develop this specification in two stages. First, in Section I, I specify three conditions that, I hold, are jointly sufficient for infringement of the RABI. Unsurprisingly, these three conditions centrally refer to bodily interference, and in the second, longer part of the chapter, I deepen the account by offering a working account of such interference. This specifies a set of conditions which I take to be jointly sufficient for an action to constitute a bodily interference.

I. A Preliminary Working Account of the Right Against Bodily Interference

What is the RABI? The answer might seem so obvious as to render the question obtuse; the RABI is a right against all and only interferences with the right-holder's body. But on reflection, things are not so clear. The RABI is not obviously infringed by *all* bodily interferences performed on individuals who possess this right. There are, I think, at least three types of case in which an intervention that plausibly constitutes a bodily interference plausibly does not infringe the RABI.

[4] This is true also in the law in some jurisdictions. See, for this point, Herring and Wall (2017).

First, it seems to me doubtful that very *minor* bodily interferences infringe the RABI (T. Douglas & Forsberg 2021). Suppose that I wave my hand near your arm, causing your arm hairs to quiver. I find it somewhat plausible that I here interfere with your body, albeit very mildly. However, it seems very doubtful that I thereby infringe your RABI. Alternatively, suppose that I turn on the lights and thereby cause photons to penetrate into your eyeballs and strike your retinas. Again, it seems to me somewhat plausible to think that I have interfered with your body in a very minor way, but not plausible to think that I have infringed your RABI.

If I indeed do not infringe your RABI in these cases, the reason, I think, is that my interference with your body is *insignificant*, in the sense that the magnitude of the interference falls below some yet-to-be-specified threshold. I will, then, restrict myself to claiming that *significant* bodily interferences—interferences whose magnitude lies above the threshold—infringe the RABI. I will not take a firm view on what determines the magnitude of an interference, nor on where the threshold for significance lies, but I take it to be plausible that the magnitude of a bodily interference will be determined by, perhaps among other things, the number of body parts that are interfered with, the importance of the interfered-with body parts, and the degree of alteration to or destruction of those body parts.

Second, a bodily interference could fail to infringe the RABI simply because it is performed by someone *against whom* the right is not held (that is, who falls under no duty that is correlative to the right). If my two-year-old son pokes his finger in my eye, he plausibly interferes with my body, but he presumably does not infringe my RABI, because he does not possess the cognitive and moral capacities necessary to fall under the correlative duty, or indeed to have any duties at all. But even people who have the cognitive and moral capacities required to have *some* duties might lack the particular duty that would correspond to the RABI in a particular case. Suppose that, unbeknownst to me, an evil scientist has connected the 'T' key on my laptop up to a machine that delivers an electric shock to the next person to press the 'up' button on a particular elevator. I then go to type my name, and thereby cause someone in a faraway elevator to experience an electric shock. Some might want to hold that, though I interfere with this person's body, I do not infringe their RABI, perhaps because I could not reasonably have been expected to foresee that my action would have those effects in virtue of which it amounts to a bodily interference and so could not have had a duty to abstain from such interference. Alternatively, suppose that I do know that the 'T' key is delivering these electric shocks, and I am, in general, cognitively and morally

competent, but right now I cannot help myself from pressing the 'T' key, because a nefarious neuroscientist is, using subliminal techniques, continually inserting a strong desire to do so into my mind. In this case, it might seem that, while I have the cognitive and moral capacities that comprise moral agency, I am being prevented from exercising those capacities. And perhaps this is enough to prevent me from bearing a duty not to interfere with the body of the elevator user in the way that I do.

I do not myself find it impossible to believe that the agent of interference *does* bear the relevant duty, so *can* infringe the RABI, in such cases. However, I think it would be controversial to assert this, so I will refrain from doing so. I will claim that a bodily interference infringes the RABI only where the perpetrator of the interference is a *well-situated moral agent*, where the 'well-situated' is meant to imply that there is nothing to prevent the agent from bearing the duty that corresponds to the RABI in the particular case in question. I will take the class 'moral agents' to include at least all adult human beings who are not severely cognitively or morally impaired, and will assume that these beings are also 'well-situated' if (i) they could reasonably be expected to foresee the consequences or other features of their conduct by virtue of which they interfere with the putative right-holder's body in the particular case, and (ii) there is nothing severely impeding the exercise of their cognitive and moral capacities.

Third, a bodily interference might not infringe the RABI because the right-holder has waived or forfeited the right in relation to that interference. Suppose that, while a patient is in a temporary state of delirium, a doctor pins him down and injects him with a calming drug. But suppose that a short time previously, while fully *compos mentis*, the patient had freely consented to being treated in this way, were he to enter a state of delirium. I take this to be a case in which the doctor interferes with the patient's body, but likely does not infringe his RABI, because, in consenting to the procedure, the patient has likely waived the right in relation to that interference.

Alternatively, suppose that, while in possession of all of their cognitive capacities and free of interference from others, a person attacks me on the street, and I respond by attempting to physically restrain them. In this case, I take it that I interfere with the person's body, but likely do not infringe their RABI, because they have likely forfeited the right as it applies to this interference.[5]

[5] We could instead think of a case like this as one in which the attacker's RABI is infringed by the defender, though justifiably so. However, I think it is more plausible to think that the RABI is forfeited, so not infringed. If the RABI were infringed, it is natural to think that the defender

I will not commit to any specific view regarding what it takes to waive or forfeit the RABI, but as the examples I have just given suggest, I remain open to the possibility that consenting to an interference may waive the right in relation to that interference, and committing a wrong may forfeit the right in relation to an interference.

I will, then, allow that interferences with the bodies of those who possess a RABI may not infringe the RABI when the interferences are insignificant, are performed by beings who are not well-situated moral agents, or are performed on individuals who have waived or forfeited the right in relation to the interference in question. But I will claim that bodily interferences *do* infringe the RABI when they are (i) significant, (ii) performed by well-situated moral agents, and (iii) performed on individuals who possess the right and have not waived or forfeited it in relation to the interference in question. These conditions—(i) through (iii)—are jointly sufficient for infringement of the right.

Some may suspect that I could do without these conditions, since they may think that it is part of the concept of bodily interference that it infringes some right over the body. Thus, in cases where there is no rights infringement, for one of the reasons outlined above, there is also no bodily interference. I acknowledge that there may be *a* concept of bodily interference on which bodily interference necessarily involves a rights infringement.[6] But I think that there is also a less normatively loaded concept of bodily interference which has no such implication: a concept on which we can perfectly well speak of, say, validly consented-to surgery as a form of bodily interference, though it infringes no bodily rights. In this book, I will be interested in the latter concept of bodily interference, and, as I will later explain, the analogous concept of mental interference.[7]

would owe the attacker compensation, or at least an apology, for having wronged them, albeit justifiably. But it seems to me that the defender owes the attacker no such thing.

[6] Jonas Hertel has suggested to me that there might also be a conception of bodily interference on which, though bodily interference does not always infringe a right, it does always *implicate* a right, in the sense that there would have been a rights infringement had the right not been waived or forfeited. I agree that 'bodily interference' may sometimes be used in this way, but again, this is not the conception of bodily interference of interest to me here.

[7] Though I do not think that bodily interference, of the type I am interested in, necessarily involves the infringement of any right, I do think it necessarily involves some agent who is interfering. That is, I am interested in an *agential* concept of interference. When I say that someone's body has been interfered with, I mean that some agent has interfered with it, not, for example, that it has been interfered with by some natural force, like a falling tree.

II. Deepening the Account: A Working Account of Bodily Interference

Central to the RABI, as I have just characterised it, is the concept of bodily interference. Yet thus far I have said almost nothing about what bodily interference is.

We generally know bodily interference, or at least significant bodily interference, when we see it. However, like the right against it, bodily interference is somewhat under-theorised. Discussions of it can be found in the literature,[8] but there is nothing approaching a standard account of bodily interference that we can simply plug in to our analysis. Hence, in this section, as in the previous one, I will be relying heavily on intuition. In particular, I will be relying on two types of intuitions: intuitions about whether it is apt to describe a particular intervention as a bodily interference, and intuitions about whether an intervention is morally wrongful. For the most part, I will rely on intuitions of the former sort. However, when these intuitions are not decisive—when it is neither clearly apt nor inapt to characterise an intervention as a bodily interference—I will sometimes advert to specifically moral intuitions. For example, when it is unclear whether an intervention can be aptly described as a bodily interference, I will take it to count in favour of classifying it as such that it seems morally equivalent to an intervention that clearly does qualify as a bodily interference.

A. Four Negative Claims

Before beginning to develop my positive account of bodily interference, let me start by offering four negative claims—four claims about what does *not* determine whether an intervention qualifies as a bodily interference.

First, an intervention need not cause net, or indeed any, harm in order to count as a bodily interference.[9] Nor need it be the case that one could reasonably expect that it will cause (net) harm. In a medical context, for example,

[8] See, for example, Earp (2019), Holm (2025), Ryberg (2020), and Wilkinson (2011). These accounts typically understand bodily interference as one or a disjunction of bodily touching, bodily invasion, and bodily alteration. For discussion, see Tesink et al. (2023).

[9] For other statements of this view, or close analogues of it, see, for example, Archard (2008, p. 27) and Ripstein (2006, pp. 235–36). Note that my claim here is consistent with thinking that (net) harmfulness typically *is* relevant to *how seriously* a bodily interference infringes the RABI. I will not take a stance on whether (net) harm is relevant in this way.

the fact that a medical or surgical procedure is 100% safe and overall in the patient's best interests, and this is clear in advance, would not prevent it from qualifying as a bodily interference.[10] Thus, if a doctor injects a vaccine into a patient, she interferes with his body, no matter how safe the vaccine is, and no matter how much the patient has to gain from it.

Second, it is also not necessary that the intervener has bad motives, or lacks good ones, for performing the intervention. The well-meaning doctor who injects a vaccine into a patient interferes with that patient's body regardless of how good her motives for injecting the vaccine may be.

Third, I believe that it is also not necessary, to interfere with someone's body, that one intends or foresees any impact on their body, or even that one could reasonably be expected to foresee any such impact.[11] Suppose I jump out of the way of a falling tree and, not noticing you or having any reason to expect you might have been there, bump into you and knock you to the ground. It seems to me that, in this case, I interfere with your body. Or suppose that, unbeknownst to me, there is an unconscious person trapped under the floorboards in my living room and that, every time I walk across the floorboards, I squash that person's body a little bit. Again, it seems to me that I interfere with that person's body. In this respect, bodily interference is, I think, like encroachment on a person's land. Suppose it turns out that part of the pavement on which I walk to work every day is—though it appears just like any other bit of public pavement—in fact someone's private property. It seems clear that I have encroached on that person's property.

Finally, fourth, it is not necessary, to interfere with a person's body, that one interferes with any particular part of that person's body. For example, that an intervention affects only some relatively peripheral or unimportant body part—such as one's little toe—would not itself prevent it from counting as a bodily interference.[12]

[10] This is plausibly why both the law and codes of medical ethics standardly require that, for a medical or surgical procedure to be permitted, it must, in typical circumstances, be both in the patient's best interests *and* validly consented to by the patient. It is the consent, not the promotion of best interests, that may prevent the procedure from infringing the right against bodily interference.

[11] Though as I acknowledged in Section I, when one could not reasonably have foreseen the features of one's conduct by virtue of which one interferes with a person's body, this may prevent one from infringing their RABI. Foreseeability may be relevant to RABI-infringement in ways that it is not relevant to bodily interference.

[12] The nature of the body part(s) affected may however be relevant, in a derivative way, to whether a bodily interference infringes the RABI. This is because it may be relevant to whether a bodily interference counts as *significant*. For example, it may be that interferences with one's heart or brain more easily count as significant than interferences with one's toenails.

B. Bodily Interference as Bodily Touching

So much for what does not determine whether an intervention counts as a bodily interference. Can we say anything about what *does* determine this? As a rough starting point, we might think that you interfere with a person's body when you touch her body, either with your own body, or with another object.

I suspect that this is indeed a sufficient condition for bodily interference. Moreover, I think it captures many of the paradigmatic cases of bodily interference, such as covertly injecting someone with a vaccine or drug, fracturing a person's skull in a violent assault, or surgically removing a person's kidney without her consent; these interventions all involve one person touching the body of another. But it misses other cases. Suppose I expose a person to electromagnetic radiation, thereby inducing radiation burns. It seems clear that I have interfered with this person's body, though my intervention may not have involved any touching.[13] Perhaps we could revise our common-sense understanding of touching so as to accommodate touching with electromagnetic radiation.[14] But there are other bodily interferences that could not be accommodated even under a highly revisionary conception of touching. Suppose Ava sucks most of the oxygen out of the room in which Bram is sleeping, thus causing Bram to suffer hypoxic damage to many of his organs. This, it seems to me, is a case of bodily interference, yet there is clearly no touching in this case.[15]

It might be thought that in these cases—causing burns through administering radiation and causing organ damage by depriving a person of oxygen—it is the presence of bodily *harm* that explains why we think there is bodily interference. However, while I think these cases likely do involve bodily harm, I do not think that this harm explains why they also involve

[13] The touching condition also misses cases of 'mere' orifice invasion. Suppose I insert a toothpick 1 cm into your nostril without touching your skin, nose hairs, or nasal mucosa. I think it is at least arguable that I have interfered with your body, though I have not touched you. However, since I am not confident that this case involves bodily interference, I am going to set it aside; I will not seek to accommodate such 'mere orifice-invasion' cases within my account. (Since my account offers only sufficient conditions for bodily interference, not necessary ones, I can remain open to the possibility that mere orifice invasion does in fact qualify as bodily interference.)

[14] For discussion of this possibility, see Cabrera et al. (2014), Fitz and Reiner (2015, p. 411), and Ford and Deshpande (2013, p. 317).

[15] In the law, too, one can interfere with something without touching it. For example, in the seventeenth-century English cases *Farmer v Hunt* and *Durant v Childe,* the defendants chased the claimants' cattle, causing them to die, but without touching them, and were found liable in trespass against property. See, for discussion, S. Douglas (2011, p. 111).

bodily interference, for notice that the bodily interference would remain even if the harm were subtracted from these cases. Suppose that, as it happens, Bram suffers from far too much oxygen in the blood, and *this* is causing organ damage. Ava sucks much of the oxygen out of the room in which Bram is sleeping, thereby dramatically lowering Bram's blood oxygen levels and correcting this damage. Even here, it seems to me that Ava interferes with Bram's body. The presence of bodily interference in this case is also supported by the intuition that Ava had reason to obtain Bram's consent prior to performing this intervention; one plausible explanation for why she had this reason, I suggest, is that she is interfering with his body and will, if she does so without his consent, infringe his RABI.

C. Bodily Interference as Bodily Alteration

How can we account for the bodily interference present in these radiation-exposure and oxygen-removal cases, including variants of them that inflict no bodily harm? I suggest that we can do so by positing that there is a second way in which one can interfere with a person's body; one can interfere with a person's body by touching their body, but one can also interfere with their body by *altering their bodily states* (Ryberg 2020, p. 76).[16] In sucking the oxygen out of the room in which Bram is sleeping, Ava interferes with Bram's body by altering his bodily states. And she does this even though she does not touch his body and does not harm him.

This raises two questions. What does it take to alter someone's bodily states? And which alterations of a person's bodily states—if not all—constitute

[16] It might be thought that, if we allow for bodily interference via bodily alteration, we can do without a separate 'touching' condition for bodily interference, for it might seem that touching a person's body always involves altering their bodily states, at least at a microscopic level. I think, however, that we should retain the touching condition. There are two reasons to do so. First, though they may be scientifically unrealistic, it is possible to imagine cases in which one touches another's body but causes no bodily alteration, even at an atomic level (suppose that the laws of physics are momentarily suspended, allowing all of the electrons, protons and neutrons in a person's body to continue on their trajectories even though some of them are touched by another object). And it seems to me that there is bodily interference in such cases. Second, bodily interferences can, I think, count as *significant* bodily interferences when they involve significant forms of touching, even when they do not produce any significant bodily alteration. Stroking a person's thigh counts as a significant bodily interference, but it is not clear that it needs to induce any significant bodily alteration. The most natural way to account for the significance of the bodily interference in this case is, I think, to allow that the thigh-stroking counts as a bodily interference by virtue of the touching that it involves, but not (or not obviously) by virtue of the bodily alteration that it produces.

bodily interferences? I will address the first of these questions in this subsection, and the second in the subsequent subsections.

The first question raises some difficult issues. To alter a person's bodily states is presumably to bring about or contribute to some alteration in her bodily states, but what is a bodily *alteration* exactly? Are we, for instance, to understand 'alteration' in *temporal* terms (as a change compared to some prior state) or in *counterfactual* terms (as a change compared to some state that would otherwise have obtained)? Another set of issues concerns the relationship between my conduct and the alteration. To alter a person's bodily states, do I, for instance, have to *cause* an alteration to them, or is it enough that I *allow* the alteration to occur? I am going to examine these and related questions in some detail, since the answers to them will be important for the account of *mental* interference that I will develop in subsequent chapters.[17]

Suppose that I can prevent a child's feet from growing by briefly subjecting them to a special sort of magnetic field. This procedure, which involves no bodily touching and causes no pain, freezes the physical development of her feet; they stay the same size, and retain exactly the same internal structure, that they had at the time I employed the procedure. It seems to me that I interfere with her body. Yet we cannot understand the interference here in temporal terms, for I cause no change to her feet across time—I simply preserve the *status quo*. This may indicate that temporal alteration is not the type of alteration that is relevant to bodily interference.

Perhaps, then, we should understand 'alteration' in counterfactual terms. Perhaps what matters is that I bring it about that her feet do not grow when they would otherwise have grown. But now consider the following case: I employ the magnetic field procedure to prevent a child's feet from growing, but had I not done so, someone else would have done the same, and at the same time. The child's feet would in any case have been prevented from growing, and in exactly the same way. Again, it seems to me that I interfere with the child's body by preventing her feet from growing. But I have not brought about any bodily state that would not otherwise have obtained. This suggests that counterfactual alteration, as understood above, is also not the type of alteration relevant to bodily interference. What type of bodily alteration *is* relevant then?

I will assume that bodily alterations should be identified by comparing the actual situation to *some* counterfactual situation, or range of such situations,

[17] Similar questions could be raised regarding 'touching'. I will not examine those, since they will not be relevant to my account of mental interference.

but not necessarily the one that would have obtained had the putatively interfering agent acted otherwise. I will refer to the situation(s) that should serve as the comparator as the 'baseline' or the 'baseline situation(s)'.

In the case that I just gave, it seems reasonably clear what the baseline is: it is a situation, or range of situations, in which *no one* performs the magnetic field procedure, or otherwise prevents the child's feet from growing, and they go on growing. Unfortunately, however, it is difficult to say what makes this the baseline.[18] In many, and perhaps all, cases, the baseline will, I take it, be the 'normal' situation(s) and the situation(s) that could have been 'expected' to come about had the agent not acted as he did. But 'normal' and 'expected' could be understood in different ways, and I do not know how to characterise them in a way that avoids all problems. So my suggestion is a rather unsatisfactory one. However, I will seek to mitigate this defect by focussing throughout on cases in which it is intuitively clear what the baseline is or in which it is at least intuitively clear that the putatively interfering agent has brought about a situation that differs from the baseline, whatever that baseline is.

One upshot of this approach is this: I will not claim that any bodily alteration has occurred when the situation brought about by the putatively interfering agent has a good claim to *be* the baseline situation, or among the set of baseline situations. Suppose that I am a water engineer asked to decide between two different filtration systems for the city's new water supply. One system will result in the water containing slightly lower levels of salt, and so in the city's residents having slightly lower blood salt levels, than if the other system had been employed. But in both cases the salt levels of the water supply and the salt levels of the city's residents will be clearly within the typical range for water supplies and for human blood, respectively. Moreover, suppose that in recent years, the salt levels in the city's water supply have fluctuated very substantially between levels much higher and much lower than those that would be consistently achieved by either of the new systems. In this case, whatever option I take, I will produce a situation that has a good claim to be among the range of baseline situations. Thus, following my approach, I would not claim that there is any bodily alteration in this case. On the other hand, were I to treat the water in such a way as to give it an extremely high salt content—much higher than had ever been witnessed in this

[18] For an excellent demonstration of some of the difficulties in answering a similar question in relation to harm (What kind of reduction in well-being is relevant to comparative harm?), see Johansson and Risberg (2019).

city or others before—then I would, I take it, clearly have brought about a bodily alteration, since it would then be clear that the situation I produce differs from the baseline.

So far, we have been discussing what it takes for an alteration in a person's bodily states to occur. But recall that we were exploring the possibility that you might interfere with another's body by *altering* their bodily states, and altering another's bodily states requires more than that their body happens to undergo some alteration—it requires also that you, the putatively interfering agent, stand in the right relationship to that alteration. To alter a person's bodily states, is it enough that you *allow* an alteration to occur, or must one make some more positive causal contribution? If the latter, what sort of positive causal contribution, exactly? And must you contribute to the alteration through performing an action, or may you do so through omission, or perhaps even just through some involuntary movement (or nonmovement) of your body?

To begin to explore these questions, consider the following case. Suppose that someone has put a poisonous plant in your bedroom—a plant prone to cause a skin disease. I know this, and I could easily remove the plant, but I elect not to do so. As a result, you develop the skin disease. I have clearly allowed the skin disease to occur, and this plausibly constitutes a deviation from the baseline (which we may suppose to be a situation, or range of situations, in which you are not exposed to the plant and do not develop the disease). So I seem to have allowed an alteration to occur. But have I *altered* your bodily states? I do not have a clear intuition on this, so I will not assert any such claim. I will claim that one individual alters the bodily states of another only where the former makes some more positive causal contribution to the alteration than is implied by 'allowing' the alteration to occur.[19]

Of course, the conceptual and normative problems raised by the distinction between allowing and more positive forms of causal contribution are legion. I cannot do anything to resolve these problems here.[20] Instead, I am simply going to help myself to the distinction, insofar as it applies to bodily alterations. However, I will seek to evade the most serious problems

[19] This is consistent with it being the case that, in failing to remove the poisonous plant, I do in fact interfere with your body in a way that is not captured by my account; recall that I am offering sufficient conditions for bodily interference, not necessary ones. Moreover, even if I do not interfere with your body, I may still wrong you in other ways, for example, by harming you.

[20] For classic discussions of some of the problems, see Bennett (1993), Foot (1984), and Quinn (1989).

by focussing on cases where it is clear that the intervening agent contributes to the alteration in a way that goes beyond merely allowing it to occur.

I am also going to accept that there may be cases in which an individual goes beyond merely allowing an alteration—she makes a positive causal contribution to it—yet still does not count as altering the other's bodily states. Suppose I plant a seed in a forest, that seed grows into a tree, and, 20 years later, a gust of wind fells that tree resulting in it striking the leg of a passing runner. Here, it seems plausible that I have not merely allowed a bodily alteration to occur, I have causally contributed to it in a more positive way. But it is not clear that I have altered the bodily states of the runner, nor indeed that I have interfered with her body. This may be because my causal contribution to the alteration is too small in magnitude. Or it may be because my causal contribution to the alteration is too causally distant; the length of the causal chain between my action and the alteration may be too great.

Consider also cases of intervening agency. Suppose I instruct Carl to pierce your ear while you are sleeping, and Carl, being keen to impress me, does this. Here it is clear that Carl alters your bodily states, but it is less clear that I do so, even though I seemingly do make a positive causal contribution to the alteration (I do so by making a positive causal contribution to Carl's action). Once again, it seems to me that my positive causal contribution to the alteration may, in this case, be too minor or too causally distant from the alteration for me to count as altering your bodily states.

I will avoid claiming that one individual has altered the bodily states of another in cases like those discussed in the previous two paragraphs. However, I will take it to be sufficient to alter a person's bodily states that one makes a *major* and *proximate* positive causal contribution to the alteration, where the 'proximate' implies that there is only a short causal distance between my contribution and the alteration. I remain neutral on what exactly it takes for a positive causal contribution to be major and proximate. In line with my strategy in relation to the other complications described in this subsection, I will simply focus throughout on cases in which the causal contribution is sufficiently great and sufficiently proximate that it would clearly be apt to say that the putatively interfering agent *altered* the other's bodily states.

A final question concerns the role played by the interfering agent's agency. To alter one of your bodily states, must I perform some *action* that causally contributes in the right way to the alteration, or is it enough that I omit to do something, or perhaps even just that some part of me, of my body, or of my mind, contributes in the right way to the alteration? Suppose the wind blows me into you as I am walking across a bridge, and the collision knocks you to

the ground. Have I altered your bodily states? I do not assume so. However, I do take it that I have altered your bodily states if I *perform an action* that makes a major and proximate positive causal contribution to the alteration. Moreover, I here and henceforth understand actions to include intentional omissions. Thus, suppose, for example, that I emerge from unconsciousness after being caught in an earthquake and I find myself lying on, and squashing, your leg. I take it that I would count as altering your bodily states if I could roll off your leg, but I intentionally do not do so.

On my account, then, it is sufficient to alter someone's bodily states that one performs an action—where that action might be an intentional omission—that makes a major and proximate positive causal contribution to an alteration in that person's bodily states.

D. Narrowing the Account

I have been suggesting that one *can* interfere with a person's body by altering their bodily states. But it does not follow that, *whenever* one alters another's bodily states, one interferes with their body. In fact, some actions that alter another person's bodily states—henceforth sometimes 'bodily influences'—seem clearly not to involve bodily interference.[21] Consider the following case:

> *Jump Reward:* Dieter wants to see how high Erin can jump. He tells Erin that he will give her £10 if she can jump high enough to touch the ceiling. Erin jumps up and touches the ceiling.

In this case, Dieter plausibly alters Erin's bodily states, as I characterised 'alters' in the previous subsection: he performs an action that seemingly makes a major and proximate positive causal contribution to an alteration in Erin's bodily states. Yet it seems clear that Dieter has *not* interfered with Erin's body.

Perhaps it could be said that Dieter has not altered Erin's bodily states because the bodily alteration is ultimately produced by an action of Erin's. This might make this case relevantly like cases of intervening agency in which *A* contributes to an alteration of *B*'s bodily states by inducing *C* to act in a way that produces the alteration; recall the case in which I ask Carl to pierce your ear while you are sleeping, and he does that. I suggested above that, in this

[21] This point is made also by Ryberg (2020, pp. 78–79).

ear-piercing case, I might not count as altering your bodily states because my positive causal contribution to the bodily alteration may be too minor or too causally distant from the alteration. Perhaps similar reasoning would apply to *Jump Reward*. This is, after all, also a case in which one agent's action produces a bodily alteration via the action of another. One difference between *Jump Reward* and the ear-piercing case is that, in *Jump Reward*, the intervening agent and the agent whose bodily states are altered are one and the same person, but it is not clear that this difference is of any significance. So it might be thought that, just as it is unclear that I alter your bodily states by getting Carl to pierce your ear, so too it is unclear that Dieter alters Erin's bodily states by offering her the £10.

There are, however, other cases in which it is clearer that one individual alters the bodily states of another, yet it remains doubtful that the former interferes with the latter's body.

Consider:

> *Lion Warning.* Fernando wants Galina to experience a large elevation in her stress hormone levels because he knows there is a lion on the loose in the neighbourhood and that she needs to be vigilant and ready to run to safety. He truthfully tells Galina about the lion and her stress hormones indeed spike.

In this case, Fernando alters the state of Galina's endocrine system, and the alteration is not clearly mediated by any other action, whether performed by Galina or a third party. But it is at best doubtful that he interferes with her body. Alternatively consider:

> *Rain Evidence.* Hachiro wants Ingrid to realise that it's probably about to rain, so he presents Ingrid with good evidence for this. For example, he truthfully tells her that his weather app shows a 100% chance of rain, and he points to the dark clouds gathering upwind on the horizon. Ingrid comes to believe that it is going to rain shortly.

Here, Hachiro has presumably altered some of Ingrid's bodily states—specifically, the neural states that underpin her beliefs about rain. And it is not clear that he has done so via any action performed by Ingrid or anyone else. (It's not clear that forming a new belief, as Ingrid does, counts as an *action* on Ingrid's part.) But again, it's doubtful that Hachiro has interfered with Ingrid's body.

I wish to avoid committing myself to the presence of bodily interference in *Lion Warning* and *Rain Evidence*. How can I do this? It might be thought that the reason there is no (clear) bodily interference in these cases is that, in them, the bodily alteration is produced through a partly mental process within the influencee. Thus, in response to these cases, we might be tempted to add something like the following further condition to our working account: the processes within the influencee via which the bodily alteration is produced are *purely physical*—they do not include mental processes, such as perception, reflection, and reasoning, within the influencee. This would leave us with the following set of jointly sufficient conditions for bodily interference: *A* interferes with *B*'s body if

> *A*'s action *X* alters a bodily state of *B*'s

and

> the mediating process—that is, the process within *B* via which *X* produces the bodily alteration—is not, and does not include, a mental process.

These conditions capture the paradigmatic cases of bodily interference via bodily alteration while avoiding the implication that *Lion Warning* and *Rain Evidence* involve bodily interference, since in those cases, the means via which the alterations are produced include mental processes within the influencee.

I think that these conditions *are* jointly sufficient for bodily interference. However, I also think that they fail to capture some forms of bodily alteration that we arguably should class as bodily interferences. Consider:

> *Flashing Lights.* Jarvis causes Kira's left foot to twitch for several minutes while she is sleeping. He does this by exposing her to lights of a particular colour, flashing at a particular frequency. Subconsciously seeing the lights through her closed eyelids causes a localised epileptic seizure in the part of Kira's brain controlling her left foot. Kira does not consciously register the flashing lights, the seizure, or the twitching of her foot.[22]

[22] The potential for flashing lights to trigger epileptic seizures is well established. More recently, evidence has begun to accumulate suggesting that flashing lights can also be used to modulate memory. See, for a review, Blanco-Duque et al. (2024).

> *Red Cabbage*. Marnie has frequently experienced vomiting episodes shortly after eating red cabbage. In fact, the red cabbage played no role in causing the nausea. Nevertheless, Marnie now experiences stomach contractions whenever she smells red cabbage. Knowing this, Lexi—who wants Marnie to experience stomach contractions—puts a pot of red cabbage in Marnie's room as she is sleeping, and Marnie's stomach indeed begins to contract. Marnie does not consciously register either the smell of red cabbage, or the stomach contractions.

In these cases, the bodily effect of the intervention is produced in part via perception, which is arguably a mental process, and in *Red Cabbage*, other mental processes (such as subconscious associative processes) are involved as well. Yet it seems to me plausible, though admittedly not obvious, that there is bodily interference in these cases. More importantly, if the concept of bodily interference is to have any moral significance—for example, if it is to serve as the basis of a moral right—it is hard to see how one could deny that there is bodily interference in these cases, for it is doubtful that there is any morally significant difference between them and comparable cases in which there is clearly bodily interference. It is, for example, hard to see any morally significant difference between these cases and cases in which the leg twitching and stomach contractions are produced through magnetically stimulating the relevant nerves using magnets placed outside the body.

How can we accommodate *Flashing Lights* and *Red Cabbage* within our working account of bodily interference? The obvious approach would be to tinker with the second of the two jointly sufficient conditions mentioned above—the condition referring to the mediating process. Here is my proposed alternative condition: a bodily influence constitutes bodily interference if the mediating process—the process within the influencee via which the alteration is produced—is *insensitive to the reasons that bear on the alteration*. Call this condition the *insensitive process condition*.

It seems to me that the insensitive process condition will allow us to explain why there is bodily interference in *Flashing Lights* and *Red Cabbage* without implying that there is also bodily interference in *Lion Warning* or *Rain Evidence*. But before I explain why I think this, I need to further describe several aspects of the condition.

E. The Insensitive Process Condition

Exactly what the insensitive process condition implies will depend on (a) how we understand 'reasons', (b) what reasons we take there to be, (c) which reasons 'bear on' a bodily alteration, and (d) what it takes for a process that produces an alteration to be sensitive to the reasons that bear on the alteration.

Consider first (a) and (b): what is a reason, and what reasons are there? I will understand reasons to be considerations (and more specifically, facts) that count for or against something.[23] The fact that Galina is at risk of a painful death is, plausibly, a reason for her to be vigilant, and thus, perhaps, to undergo a change in stress hormone levels that would facilitate such vigilance. The fact that there are dark clouds on the horizon upwind is plausibly a reason for Ingrid to form the belief that it will rain shortly, and thus, perhaps, to undergo the neural changes that underpin this belief-change. I will not commit to any highly controversial stance on what kinds of facts can count as reasons, and in line with my general approach, I will try to be ecumenical. In particular, I will allow that a person may have both preference-based reasons (fundamental reasons that are facts about the person's preferences, and derivative reasons that derive from such fundamental reasons) and non-preference-based reasons (fundamental reasons that are not facts about the person's preferences, and derivative reasons that derive from such fundamental reasons). Thus, consider again *Lion Warning*. I suggested that in this case, the fact that Galina is at risk of a painful death might be a reason for Galina to be vigilant. This reason might derive from a more fundamental reason that is a fact about Galina's preferences—such as the fact that Galina *prefers* not to die a painful death. In that case, it will count as a preference-based reason. Alternatively, however, it might derive from a fundamental reason that is not a fact about her preferences, like the fact that a painful death would be objectively bad for her. In that case, the reason will be non-preference-based. I will refer to the kinds of preferences that can ground preference-based reasons as *authoritative preferences*. I will attempt to remain neutral on precisely what determines whether a preference qualifies as authoritative, but considerations that would plausibly be relevant to this

[23] In a deviation from standard usage, I will allow that a fact is a reason even when it counts for or against something only in the relative sense that, if the fact did not obtain, that would count (in an absolute sense) for or against that thing. Thus, for example, I will allow that the fact that the element is not hot is a reason not to move your hand off it because, if it were hot, that would count absolutely in favour of moving your hand off it.

would include whether the preference is firmly and consistently held, and whether it is reflectively endorsed.

Consider next (c): which reasons are the reasons that 'bear on' an alteration? Here, I mean to include both reasons for or against the influencee undergoing or bringing about the alteration.[24] Thus, in *Lion Warning*, the fact that she is at risk of a painful death is plausibly a reason that bears on the alteration to Galina's stress hormone levels, for it is plausibly a reason for her to undergo or bring about this alteration (assuming that the alteration would help her to be more vigilant). But so too is the fact, if it is a fact, that Galina has a heart condition that could be triggered by a spike in her stress hormones, since this is a reason for her *not* to undergo or bring about the hormone spike. Note also that I do not require that the reasons that bear on an alteration must be epistemically accessible to the individual who would undergo the alteration. The fact that she is at risk of a painful death can count as a reason that bears on Galina's change in stress hormone levels whether or not Galina has any way of discovering this fact.[25]

Finally, consider (d): what does it take for a process that produces an alteration to be sensitive to the reasons that bear on the alteration? I will have a lot more to say about reasons-sensitivity—although in relation to mental rather than bodily alterations—in Chapter 6. For the moment, I leave reasons-sensitivity largely as an intuitive matter. I will, however, make two significant assumptions regarding reasons-sensitivity.

First, I assume that reasons-sensitivity does *not* require the conscious or deliberate assessment of reasons. Thus, for example, intuitive processes can, on my view, be reasons-sensitive. What makes a process reasons-sensitive is that it somehow *tracks* the reasons that there are. It need not track those reasons in a conscious or deliberate way.

Second, I will assume that, at least in typical cases, a simple counterfactual test will provide us with evidence (though not necessarily decisive evidence) on whether the mediating process was reasons-sensitive:

[24] If you think that having a reason to ϕ requires being able to ϕ, being able not to ϕ, or being able to determine whether one ϕs, please interpret the 'reasons for or against the influencee undergoing or bringing about the alteration' as referring to the reasons that the influencee would have had for or against these things *if* they had had the relevant abilities.

[25] Some authors distinguish between 'a person having a reason' and 'there being a reason for that person', taking the former, but not the latter, to imply that the reason is somehow accessible to the person (see, e.g. Kaiserman 2021, p. 689). I do not distinguish these things, and do not take 'having a reason' to imply that the reason is accessible; I can have a reason of which I am not, and could not become, aware.

> If it is the case that either (i) the alteration would have been more likely to occur, other things being equal, had some actually-present reason for the alteration been weaker or absent, or (ii) the alteration would have been *less* likely to occur, other things being equal, had some actually-present reason *against* the alteration been weaker or absent, then the process via which the alteration was produced was probably sensitive to reasons. On the other hand, if, for all actually present reasons bearing on the alteration, neither (i) nor (ii) holds, the process was probably not sensitive to reasons.

How do the cases that we have discussed fare against the insensitive process condition? Consider first *Flashing Lights*, in which Jarvis causes Kira's leg to twitch while she sleeps by exposing her to flashing lights. In this case, the process within Kira via which Jarvis' flashing lights procedure causes Kira's leg to twitch is presumably insensitive to reasons. Suppose that, as it happens, Kira does have a reason to twitch her leg; there is a spider about to bite her toe and twitching her leg will scare it away. It is very doubtful that the process mediating the flashing lights intervention is sensitive to this reason, or, I think, any other reason bearing on the alteration. Presumably, the intervention would have been just as likely to produce the twitching had Kira lacked this reason, and the same goes for other actually present reasons we might identify. Similar comments apply to *Red Cabbage*. It is doubtful that the conditioned process within Marnie via which the cabbage smell causes her stomach muscles to contract is sensitive to any reasons that she actually has to or not to undergo or bring about the contractions.

But the mediating process in *Lion Warning*—the process within Galina via which Fernando's warning alters her hormone levels—plausibly *is* sensitive to the reasons that bear on the alteration. For example, suppose that Galina prefers not to die and prefers not to experience pain, and that these preferences are authoritative ones. The fact that she has these preferences gives her reasons to be vigilant about the lion, and, on the assumption that the stress hormone spike helps her to be more vigilant, plausibly gives her reasons to undergo or bring about the increase in her stress hormone levels. Moreover, the mediating process is plausibly sensitive to these reasons. Had she been indifferent to whether she died or experienced pain, it is less likely that she would have experienced the stress hormone spike as a result of Fernando's warning.

Likewise, in *Rain Evidence*, the process within Ingrid via which Hachiro's presentation of evidence concerning the likelihood of rain affects her neural states is plausibly sensitive to reasons. Recall that one of the pieces of evidence

that Hachiro made salient was the fact that there were dark clouds gathering on the horizon upwind. This fact plausibly constitutes a reason for Ingrid to undergo or bring about the change in beliefs—and neural states—that she in fact underwent. And the process within Ingrid leading up to these changes was plausibly sensitive to this reason; had the wind been blowing in the other direction, there is a chance that Ingrid would have noticed that, and in consequence would not have undergone the change of her beliefs or mental states.

I suggest, then, that the insensitive process condition can distinguish the interventions in *Flashing Lights* and *Red Cabbage*, which plausibly meet the condition, from those in *Lion Warning* and *Rain Evidence*, which plausibly do not. This condition allows us to include *Flashing Lights* and *Red Cabbage* within our working account of bodily interference, while leaving *Lion Warning* and *Rain Evidence* outside it. And this, I think, counts in favour of accepting the condition.

I do not claim that the insensitive process condition is beyond reproach. Perhaps, in the end, it cannot in fact yield the intuitively plausible verdicts that I have just suggested it can yield. Or perhaps it yields implausible verdicts in other cases. Still, I want to tentatively endorse the insensitive process condition as at least a plausible candidate for the role of determining which bodily influences—actions that alter a person's mental states—constitute bodily interferences. I will thus adopt it as part of my working account of bodily interference.[26]

III. Conclusion

In this chapter, I have offered a working account of the RABI. I first offered a preliminary account which consisted in three jointly sufficient conditions for infringement of the right. I claimed that it is sufficient, for *A* to infringe *B*'s RABI, that *A* is a well-situated moral agent, *A* significantly interferes with *B*'s

[26] It might be thought that the insensitive process condition ensures that no *consensual* influence on a person's body could count as a bodily interference. Suppose a surgeon removes my appendix, but does so only because I validly consented to this. In this case, whether my appendix is removed will be sensitive to my authoritative preferences, and thus preference-based reasons, since those preferences will inform whether or not I validly consent to the intervention. Thus, it may seem, the current account implies that there can be no bodily interference in such a case. This picture is, however, not quite right. Note that in the case of the surgeon, what is sensitive to reasons is not the mediating process—the process within me via which the surgeon's action produces the alteration. Rather, what is sensitive to reasons is the process that leads up to the surgeon's action. The reasons-sensitivity is upstream, not downstream, of the influencing action. Thus, a consensual surgical procedure will still qualify as bodily interference, on my account.

body, and *B* possesses the RABI and has not waived or forfeited it in relation to the interference in question. I then deepened this account by offering a working account of bodily interference. On this account, there are two ways of interfering with a person's body. *A* interferes with *B*'s body if:

(i) *A* touches *B*'s body with *A*'s own body, or with another object, or
(ii) *A*'s action *X* alters a bodily state of *B*'s, and the mediating process—that is, the process within *B* via which *X* produces the alteration—is insensitive to the reasons that bear on the alteration.

In the next chapter, I will develop an analogous account of the RAMI, which I will then further specify in Chapter 6.

3

A Working Account of the Right Against Mental Interference

On my working account of the right against bodily interference (RABI), the right protects against significant interferences, by well-situated moral agents, with the bodies of those who possess the right and have not waived or forfeited it in relation to the interference in question. And bodily interference consists in either

(i) touching the right-holder's body with one's own body or with another object, or
(ii) performing an action that alters the right-holder's bodily states via a mediating process, in the right-holder, that is insensitive to the reasons that bear on the alteration.

With this account of the RABI in hand, let us now move on to the right against *mental* interference (RAMI). Again, we may wonder: what is this right? In this chapter, I will answer this question. In Section I, I offer a preliminary working account of the RAMI. This, like the preliminary account of the RABI offered in the previous chapter, specifies three conditions that I take to be jointly sufficient for infringement of the right. Next, in Section II, I deepen the account by supplementing it with a working account of mental interference. This account offers two jointly sufficient conditions for an action to qualify as a mental interference. Finally, in Section III, I distinguish the RAMI, as I have characterised it, from two cognate rights that have made appearances in the recent literature: the right against mental manipulation, and the right to freedom of thought.

I. A Preliminary Working Account of the Right Against Mental Interference

I will assume that the RAMI closely mirrors the RABI. Thus, I will allow that, though the RAMI protects against mental interferences, it may not protect

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0003

against *all* such interferences.[1] More specifically, I will assume that there are three kinds of case, paralleling those identified for bodily interference, in which mental interference may not infringe the RAMI.

First, there may be cases in which the mental interference is not significant—the magnitude of the interference falls below some threshold. There are difficult questions concerning what determines the magnitude of a mental interference, and the location of the threshold for significance. For the moment, I will not take a stance on these matters; I return to consider this issue in Chapter 6.

Second, as with bodily interference, I take it that there may be cases in which a mental interference does not infringe the RAMI because it is performed by someone against whom the right is not held. As with the RABI, I will assume that the right is held against at least all well-situated moral agents, where the 'well-situated' is meant to imply that there is nothing to prevent the agent from bearing the duty that corresponds to the RAMI in the particular case in question. I will take the class 'moral agents' to include at least all adult human beings who are not severely cognitively or morally impaired, and will assume that these beings are also 'well-situated' if (i) they could reasonably be expected to foresee the consequences or other features of their conduct by virtue of which they interfere with the putative right-holder's mind in a particular case, and (ii) there is nothing severely impeding their exercise of their cognitive and moral capacities.

Third, and again as with bodily interference, I will assume that there may be cases in which a mental interference fails to infringe the right against it because the right-holder has waived or forfeited the right in relation to the interference in question, for example, by consenting to it or, perhaps, by committing a serious wrong.[2]

On my working account, then, the RAMI is a right that protects against (i.e. is infringed by) significant mental interferences, perpetrated by well-situated moral agents, and perpetrated against individuals who possess the

[1] As with bodily interference, I suspect there is a conception of mental interference on which mental interferences necessarily infringe a right, but I will hold to a less normatively loaded conception on which mental interferences need not involve rights infringements.

[2] I do not claim that it is always possible to prevent a mental interference from infringing one's RAMI by waiving the right. For example, it may be that one could never waive the RAMI in relation to an intervention that would result in another person comprehensively and irreversibly controlling your mind.

right and have not waived or forfeited it in relation to the interference in question.

II. Deepening the Account: A Working Account of Mental Interference

My next task is to offer an account of mental interference. This task is in one way more difficult than the analogous task in relation to bodily interference, since our—or at least, my—intuitions are less clear. However, in another respect, it is easier, since we can take, as a model, the account of bodily interference that I have already offered. In what follows, I will set out a schematic account of mental interference that is intended to (i) capture those intuitive verdicts that I *do* feel confident in making about what does, and what does not, constitute mental interference,[3] but also (ii) closely parallels the account I have already offered of bodily interference.

Thus, for example, I will adopt analogues of the four negative claims that I made about bodily interference in the previous chapter. I will take it that it is not necessary, for *A* to interfere with *B*'s mind, that

(i) *A*'s action causes net or any harm to *B*, or can reasonably be expected to cause such harm;
(ii) *A* has bad motives, or lacks good ones, for performing the action;
(iii) *A* intends or foresees any impact on *B*'s mind, or could reasonably be expected to foresee such an impact (mental interference can be unintended, unforeseen, and unforeseeable); or
(iv) *A*'s action impacts on any particular part of *B*'s mind.[4]

What, more positively, can we say to characterise mental interference? In the previous chapter, I proposed that there are two ways in which one

[3] I am not highly confident in these intuitions and suspect that some readers will not share them. Readers who have no, or different, intuitions about the cases I discuss can read 'mental interference' as a term of art, and my account of it as stipulative and perhaps revisionary. My core arguments for the right against mental interference, offered in the next two chapters, will not be affected by thinking of mental interference in this way.

[4] As with the RABI, the nature of the mental state(s) affected may, however, be relevant, in a derivative way, to whether a bodily interference infringes the RABI. This is because it may be relevant to whether a bodily interference counts as *significant*.

can interfere with someone's body: by touching it, or by altering it via a reasons-insensitive mediating process. With the mind there is, I think, no straightforward analogue of bodily interference via touching. Indeed, except on physicalist views which equate the mind with a physical object, such as part of the brain, it is difficult to make literal sense of the idea of touching the mind. Physicalists *can* make sense of this idea. They can, for example, claim that one literally touches a person's mind when one touches whatever parts of the brain (partly) constitute the mind. But I do not wish to include, within my account of mental interference, interferences that are possible only if physicalism holds. That would be at odds with my commitment to developing a metaphysically ecumenical account of the RAMI. (Recall that I am offering only sufficient, not necessary, conditions for infringing the RAMI. Thus, what I say here is compatible with there being some way of infringing the RAMI that is not covered by my account, and that is akin to touching.)

While there is not clearly a mental analogue of bodily touching, there *is* a clear mental analogue of bodily *alteration*; we can perfectly aptly speak of altering (or influencing) a person's mental states. Let us then take, as our starting point, the thought that you interfere with a person's mind when you alter the person's mental states, or alter them in a certain kind of way.[5]

As with bodily alterations, I will take it as sufficient to alter a person's mental states that one performs an action (which may be an intentional omission) that makes a major and proximate causal contribution to the alteration in her mental states. And 'alteration' is again to be understood in counterfactual terms, though I will again not commit to any particular view about which counterfactual situation(s) should serve as the comparator, or, as I will say, the 'baseline'. I will focus on cases in which it is intuitively clear what the baseline is, or at least intuitively clear that the putatively interfering agent has indeed produced a deviation from the baseline.

Positing that one can interfere with someone's mind by altering their mental states allows us to accommodate many of the intuitively central cases of mental interference, such as the administration of psychoactive drugs or electrical brain stimulation to control a person's mood or motivations.

[5] For a similar suggestion, see Ryberg (2020, p. 85), who begins his discussion of the 'right to mental non-interference' by considering the possibility that the right might protect against any intervention that 'non-consensually influences conative, cognitive, and affective events of our inner life'.

However, again as with bodily interference, we need to acknowledge that altering another's mental states is insufficient for mental interference; not all actions that cause mental alterations involve interference with the mind. We would not say that I interfere with your mind if I give you a persuasive argument and thereby cause you to change some of your beliefs. Nor, I think, would we say that I interfere with your mind if I point to the piece of litter that you have dropped and thereby induce in you a mild pang of guilt.[6] On the other hand, we surely *would* say that I interfere with your mind if I hypnotise you without your consent, or induce you to panic through a programme of subliminal imagery.

In which cases, exactly, do actions that alter others' mental states—henceforth mental influences—qualify as mental *interferences*? Taking the lead from the account of bodily interference offered in the previous section, I propose that they do so at least when the mediating process—the process within the influencee via which the influencing action produces the alteration—is insensitive to the reasons that bear on the alteration. Once again, the 'reasons that bear on' an alteration are the reasons for or against the influencee undergoing or bringing about the alteration.[7] And I will continue to remain open to the existence of both preference-based reasons and non-preference-based reasons.

What non-preference-based reasons could bear on a mental alteration? I want to be open-minded about this, and to avoid committing to any fixed list. But in general, I think we can distinguish three types of reason that are likely to belong on such a list. First, there may be fit-based reasons that bear on a mental alteration; such as reasons to bring about or undergo a mental alteration because this will (likely) result in one's mental states better fitting the world. Thus, for example, one might have reasons to bring it about that one believes that p because p is true, and reasons to bring it about that one desires that q because q is desirable. Second, there may be reasons that are based on the noninstrumental value of the resulting mental state. Thus, one might have reasons to bring it about that one feels warm, because warmth is

[6] For a similar point, see Ryberg (2020, p. 85).

[7] As in the previous chapter, if you think that having a reason to ϕ requires being able to ϕ, or able not to ϕ, or being able to determine whether one ϕs, then please again interpret the 'reasons for or against the influencee undergoing or bringing about the alteration' as referring to the reasons that the influencee would have had for or against these things *if* they had had the relevant abilities. Further—and again, as in the previous chapter—I will allow that the reasons that bear on the alteration may count for or against the alteration only in a relative sense. For example, a fact may count for the alteration only in the sense that, if the fact did not obtain, that would count absolutely against it.

a pleasurable feeling to have, and pleasure is noninstrumentally good. And third, there may be reasons to bring about or undergo a mental alteration that are based on the instrumental value of the resulting mental state. For example, one might have reasons to form the belief that one is a good parent because doing so is likely to boost one's confidence and thereby make one a better parent.

I will have much more to say about the implications of my account of mental interference later—especially in Chapters 6 and 7. But to briefly illustrate the account, consider two different ways in which I might cause someone to believe that it will rain shortly. In the first case, I induce this belief by pointing to the gathering dark clouds on the horizon upwind, and the person recognises these as evidence that it will shortly rain. In the second case, I produce the belief simply by electrically stimulating whatever neural circuits underpin it. In the first case, I influence the person's beliefs via a process within her that is somewhat sensitive to her reasons to undergo or bring about the alteration in her beliefs (in this case, those reasons include the evidence that it will rain). In the second case, I do not.[8]

Alternatively, consider two different ways in which I might cause someone to *prefer* that it will rain tomorrow. In the first case, I induce this preference by truthfully informing the person that the rain will help his roses to bloom earlier, which is something that he also prefers. In the second case, I induce the preference simply by electrically stimulating whatever neural circuits underpin it. In the first case, I alter the person's preferences via a process that is somewhat sensitive to at least one of his reasons to undergo or bring about the alteration (in this case, a fact about his further, and I take it authoritative, preferences). In the second case, I do not.

A comment is in order here on how I understand 'mental states'. I understand this term broadly, so as to include all the states standardly characterised as mental states, including beliefs, desires, preferences, intentions, emotions, feelings, and sensations. Less standardly, I also include abilities and dispositions to or not to form such states, such as the ability to infer a justified belief from the available evidence, and the disposition to feel rage when mildly

[8] In fact, it is *possible* that an electrical brain stimulation procedure could operate via a process that happens to be sensitive to the influencee's reasons. I will briefly consider cases like this in Chapter 7. However, as I have specified the case here, there is no clear positive reason to believe that the electrical processes harnessed by this procedure would be reasons-sensitive.

provoked. My reason for classifying such abilities and dispositions as mental states can be illustrated by way of the following case:

> *Cognitive Enhancement.* Sylvie wants to get her partner, Tristan, to believe that human actions are causing climate change. She thinks—correctly, as it turns out—that Tristan would believe in anthropogenic climate change if he were just a little better at reasoning, since he would then be better at seeing the fallacies and inconsistencies in the climate-change-denying arguments with which he is bombarded online. Sylvie begins slipping a drug known to enhance reasoning ability into Tristan's coffee each morning. And this has the desired effect. Tristan does increasingly see the flaws in the arguments with which he is presented, and after a year of receiving the drug, he has become a firm believer in anthropogenic climate change.

I take it that this is a case of mental interference. Yet it might seem that the process within Tristan via which Sylvie's intervention alters his beliefs is highly reasons-sensitive. He comes to believe in anthropogenic climate change precisely by becoming better at appreciating the reasons to believe in it. Thus, Sylvie's influence on Tristan's beliefs will not be classified as a mental interference by my account.

However, Sylvie's action does not *only* alter Tristan's beliefs concerning climate change. It alters other mental states too—for example, those that figure in the reasoning processes that he engages in as a result of receiving the drug. My account would be able to classify this case as a case of mental interference if we could identify some mental alteration in Tristan that *is* produced via a reasons-insensitive mediating process—perhaps some alteration in Tristan's mental states that is produced directly by the neurochemical action of the drug, without the intermediation of Tristan's reasoning. My account would classify Sylvie's influence on *that* mental state as a mental interference. The difficulty is that, if we understand mental states as they are standardly understood—as comprising beliefs, desires, intentions, and so on—it is not clear that the drug produces any such alteration. The drug may affect Tristan's (standardly understood) mental states *only* by affecting his ability or disposition to reason in certain ways, and thereby affecting the outputs of his reasoning.

This difficulty could be avoided if we could count Tristan's *dispositions* or *abilities* with respect to his beliefs—such as his ability or disposition to form justified beliefs concerning climate change—as themselves mental states for the purposes of establishing mental interference. For there must be some alteration in such an ability or disposition that is produced by the

drug, and that explains, rather than being an upshot of, the drug's effect on Tristan's reasoning. *This* alteration will, plausibly, be produced via a reasons-insensitive neurochemical process in Tristan.

To accommodate cases like *Cognitive Enhancement* within my account of mental interference, I allow that alterations to mental dispositions or abilities can qualify as mental interferences. And to avoid the need to specify this on every occasion, I will henceforth use the term 'mental state' so as to include such dispositions and abilities. This is a deviation from standard usage, but it is a deviation that will save me a lot of awkward wording.

III. Distinguishing the Right Against Mental Interference from Cognate Mental Rights

I have now stated my working accounts of mental interference, and of the right against it. On my account of mental interference, *A* interferes with *B*'s mind if

> *A*'s action *X* alters a mental state of *B*'s (call this the *agential alteration condition*)

and

> the mediating process—the process within *B* via which *X* produces the alteration—is insensitive to the reasons that bear on the alteration (call this the *insensitive process condition*).

Further, *A* infringes *B*'s RAMI if

> *A* is a well-situated moral agent (call this the *duty-bearer condition*)

and

> *A*'s interference with *B*'s mind is significant (call this the *significance condition*)

and

> *B* possesses the RAMI and has not waived or forfeited it in relation to the interference in question (call this the *nonliability condition*).

In this penultimate section, I offer some comments on how the RAMI, as I have just characterised it, relates to two cognate mental rights that we have sometimes been thought to possess, and that might be thought to overlap significantly with the RAMI: the moral right against mental manipulation, and the moral right to freedom of thought.

A. The Right Against Mental Manipulation

In Chapter 1, I cited a number of authors who have recently defended or entertained the view that we enjoy a moral or legal right against mental interference, or something close to such a right. The 'or something close' is doing some work here; a number of the authors that I cited characterise the scope of the right they defend not in terms of mental *interference*, but in terms of some other category of influence—in each case, a category that, I assumed, might substantially overlap with that of mental interference. In most cases, that other category was mental *manipulation* (Alegre 2017, p. 224; Bublitz 2014, p. 3; Bublitz & Merkel 2014, p. 58; McCarthy-Jones 2019; Mendlow 2018, p. 2342). We should thus consider how a right against mental manipulation might relate to the RAMI that I have outlined in this chapter. More specifically, we should ask: how, if at all, does the RAMI differ from the putative moral right against mental manipulation?

The short answer is that it is difficult to say. The concept of manipulation is famously problematic; definitions of manipulation vary greatly, as do the practices that have been described as manipulative. Jennifer Blumenthal-Barby notes that

> [t]he following practices have all been described by philosophers as manipulation: incentivizing, offering, increasing options, decreasing options, tricking, using (resistible) threats of punishment, managing information, presenting information in a way that leads to predictable inferences, deceiving, lying, making a false promise, withholding information or options, slanting information, providing irrelevant inputs or crowding out relevant inputs, exaggerating information in a misleading way, using misleading packaging or misleading images, creating impressions by imagery, using loaded language, trading on fear, subliminal suggestion, insinuations, flattery, guilt, appealing to emotional weaknesses or needs, initiating psychological processes that are difficult to reverse or that lead to predictable behaviors or decisions (e.g., the tendency to continue with

> an active decision even after becoming aware that it is more costly than originally thought, or the tendency to view an option as more desirable when shown its contrast), browbeating or otherwise wearing the person. (2014, pp. 121–22)[9]

To this list I would add a further technique: the precise control of a person's mental states and/or actions via hypnosis or neurotechnological means, which is a staple of the science-fiction 'manipulation' cases commonly discussed in the literature on moral responsibility and free will.[10]

It is doubtful that those positing a right against mental manipulation are thinking that *all* of the techniques just listed would be covered by a right against mental manipulation. But none of these authors have specified in full detail which techniques they do take the right to cover.[11] This prevents me from saying anything firm about how that right relates to the RAMI. However, let me offer some speculative comments on two questions: Are there infringements of the RAMI, on my account thereof, that would not be infringements of a right against mental manipulation? And are there infringements of the right against mental manipulation that would not meet my conditions for infringing the RAMI?

I believe that the answer to the first question is likely to be 'yes'. On most accounts of manipulation, at least some of the mental or behavioural effects of a mental influence must be *intended* by the influencer for the influence to count as manipulation. This is not so for mental interference, as I have understood it; mental interferences, and indeed infringements of the RAMI, need not involve any intentional impact on the influencee's mental states. To meet my sufficient conditions for RAMI-infringement, the mental interference does need to be foreseeable, in the sense that one can reasonably be expected to foresee the consequences or other features of one's conduct in virtue of which one interferes with the right-holder's mind. If the interference is not foreseeable in this way, the agent will not count as 'well-situated', and the duty-bearer condition will not be met. However, such foreseeability is compatible with none of the mental effects of the influencing action being intended by the influencer. Thus, one could infringe someone's RAMI without

[9] See also Wilkinson (2013, p. 344).
[10] See, for example, Mele (2016, pp. 72–74) and Pereboom (2016, p. 77).
[11] Gabriel S. Mendlow does refer specifically to '*forcible* manipulation' (2018, p. 2346, my italics).

infringing their right against mental manipulation simply because one didn't intend any of the mental effects of one's action.

Further, on some influential accounts, manipulation necessarily has an *adverse* mental effect, whereas this is not so for mental interferences or RAMI-infringements. For example, on Robert Noggle's (2020, 2025) view, attempted manipulation aims to induce some mental error, such as a mistaken belief or unfitting emotion, in the manipulee, and successful manipulation in fact induces such an error. By contrast, mental interferences, as I am understanding them, need not induce any such error. Mental interference is characterised by the means via which a mental alteration is produced, not by the nature of the influencee's resulting mental states (including mental dispositions and abilities). Thus, it is possible to conceive of a mental interference that both aims to leave and in fact leaves the influencee with an unimpeachable set of mental states (including dispositions and abilities). Such a mental interference would potentially infringe the RAMI, but would not, on Noggle's account of manipulation, infringe a right against mental manipulation.

Consider next the second question: are there infringements of the right against mental manipulation that would not meet my conditions for infringing the RAMI? Again, I believe that the answer to this question is likely to be 'yes'. It may be true that a certain subspecies of mental manipulation would typically also count as mental interference. It is commonly held that one form of manipulation consists in intentionally influencing someone in a way that (intentionally) bypasses their rationality.[12] 'Bypassing rationality' is open to multiple interpretations but on one interpretation it is, I think, synonymous or coextensive with 'influencing via a reasons-insensitive process in the influencee'. On this interpretation, mental manipulations of this variety would qualify as mental interferences, and so would potentially infringe the RAMI. However, those who take bypassing rationality to be characteristic of one type of manipulation typically allow that there are other types of manipulation as well, including types that can operate via reasons-sensitive processes. For example, the category 'manipulation' is almost always taken to include influences that subvert, pervert, or corrupt a person's rationality, even where those influences operate via reasons-sensitive processes (Noggle 2020; Wilkinson 2013, p. 345; Wood 2014).[13] Thus, for instance, selectively presenting evidence for vaccine safety in a way that causes

[12] See, for example, Gorin (2014a).
[13] For discussion, see Gorin (2014a, 2014b).

people to overweight the risks of vaccination would, on many views, constitute manipulation, but, if it operates via a reasons-sensitive process in the influencee—an explicit or implicit balancing of reasons, say—it would not be captured by my sufficient conditions for mental interference, since it would not meet the insensitive process condition. It counts as manipulation not by virtue of the process by which it operates, but by virtue of its mental effects.

Indeed, a right against mental manipulation might also cover mental influences that neither operate via reasons-insensitive processes nor produce any mental error nor subvert, pervert, or corrupt the influencee's rationality. Several authors have argued that an influence can count as manipulative solely by virtue of the motives of the manipulator. Thus, for example, Moti Gorin identifies a class of manipulation that is characterised by the manipulator's indifference to whether she is influencing by giving good reasons (Gorin 2014b, pp. 92–94), and Michael Klenk argues that manipulation is not characterised by any particular means of influence or outcome but by the manipulator's negligent failure to select the mode of influence that she employed on the basis that it would reveal reasons (Klenk 2021). On these views, influences can qualify as manipulation even when they operate via reasons-sensitive processes in the influencee (and so would not meet my sufficient conditions for mental interference) and also induce no mental error or irrationality. What matters is that the influencer is indifferent or negligent in the relevant way.

I suspect, then, that a right against mental manipulation would cover at least some mental influences that would not meet my conditions for RAMI-infringement, and that my conditions for RAMI-infringement would cover at least some mental influences that would infringe a right against mental manipulation. However, this can be no more than a suspicion, since the putative right against mental manipulation awaits full description.

B. The Moral Right to Freedom of Thought

A legal right to freedom of thought is firmly established in human rights law.[14] The analogous moral right is less well established. Nevertheless, it is more

[14] This right is, for example, asserted by Article 9 of the European Convention on Human Rights (ECHR) and Article 18 of the International Covenant on Civil and Political Rights. For recent discussions, see, for example, Blitz and Bublitz (2021), Ligthart, Ienca et al. (2023), and McCarthy-Jones (2019).

familiar than the RAMI, perhaps in large part due to its well-known defence by John Stuart Mill (1909, esp. pp. 22–23). Moreover, the right to freedom of thought, whether in its moral or legal form, is sometimes thought to include a right against certain types of intervention into the mind (typically alongside a right to mental privacy, a right against punishment for one's thoughts, and in some cases a positive right to sufficient inputs into one's thought—e.g. in the form of vigorous public discussion).[15] We should thus consider how the RAMI that I am proposing relates to this somewhat more familiar right, or, at least, the part of it that protects against interventions into the mind.

Here, we face similar difficulties to those discussed in relation to mental manipulation in the previous section; I am not aware of any full and precise specification of the class of interventions into the mind that are prohibited by the right to freedom of thought. Moreover, there is great diversity in the terms that are used to quickly pick out the relevant class. Some refer to indoctrination (Bublitz 2014; Murdoch 2007; Vermeulen & van Roosmalen 2018), some to brainwashing (Bublitz 2014; Vermeulen & van Roosmalen 2018), some to thought or mind control (McCarthy-Jones 2019), some to thought or mind manipulation (Alegre 2017; Bublitz 2014; Bublitz & Merkel 2014; Farahany 2023; McCarthy-Jones 2019; Mendlow 2018), some to intrusions in or restrictions of thought (Blitz 2010), and some to interference with thought (McCarthy-Jones 2019).

Nevertheless, there are, I think, two respects in which the RAMI is likely to differ from the part of the right to freedom of thought that protects against interventions on the mind. First, whereas the right to freedom of thought is often understood as protecting only certain kinds of mental states—typically 'thoughts', though Mill also includes 'feelings'—the RAMI that I have characterised covers all mental states.[16] It is, for instance, doubtful that perceptual states are, absent effects on other mental states, covered by the right to freedom of thought. They are, after all, neither 'thoughts' nor 'feelings' on the common-sense understandings of those terms. However, perceptual states are covered by the RAMI. Similar thoughts apply to mental abilities and mental dispositions, which I have included within the category of mental

[15] Several recent discussions of the right to freedom of thought suggest that it includes a right against certain types of intervention into the mind (Alegre 2017; Blitz 2010; Bublitz 2014; Ligthart, Ienca et al. 2023; McCarthy-Jones 2019; Murdoch 2007; Vermeulen & van Roosmalen 2018).

[16] A possible exception here would be mental states that are so peripheral or unimportant that even their elimination or radical alteration could not count as a significant mental interference.

states, and thus within the scope of the RAMI, but which are neither thoughts nor feelings in the usual senses of those terms.

Second, whereas the RAMI covers interferences with our mental states, the right to freedom of thought plausibly covers only interferences with our *freedom* with respect to our mental states. These are, I think, distinct, though overlapping, categories. It is plausible that one can interfere with someone's mental states without interfering with their freedom with respect to those mental states. This might be the case, I think, if one's interference was very easy for the influencee to avoid or resist. Suppose I enhance your desire to drink Fanta by setting your television to display subliminal images of beautiful people drinking Fanta. But suppose you knew about this plan in advance, and you could easily have avoided having your desires altered simply by flicking off a 'subliminal imagery' switch on the television. However, you never got around to thinking about whether to flick this switch. So you didn't, and ended up having your desires enhanced by the subliminal images. It seems to me that, in this case, I clearly interfere with your desire to drink Fanta, but I do not obviously interfere with your freedom with respect to that desire. You were arguably free to maintain your *status quo ante* desire; you just didn't exercise that freedom.[17]

Alternatively, consider a case in which a neuroscientist nonconsensually implants in your brain a chip that is set up to detect if your desire for Fanta ever wanes, and, if this occurs, stimulate the relevant brain regions to boost that desire. Suppose, however, that your desire never wanes, so the brain chip's stimulation mechanism is never activated. This, it seems to me, is a case in which you have lost a certain kind of freedom with respect to your desire for Fanta. That freedom has been interfered with. However, the mental state which this freedom concerns—your desire for Fanta—has not itself been interfered with. Thus, my conditions for RAMI-infringement will be met in this case only if there is some other mental state that *has* been interfered with. I think it is unclear whether this must be so. (Perhaps we could say that the neuroscientist has interfered with your mind by interfering with your ability or disposition to alter your desire for Fanta. But whether this is plausible will depend, I think, on how we understand mental abilities and dispositions, and specifically on whether these things can be affected by external constraints. I do not want to take a stance on these issues here. My

[17] In this case, I have interfered with at least one freedom: your freedom to *both* maintain your *status quo ante* desire *and* not flick the switch. However, I think it is doubtful that this freedom is an aspect of your freedom of thought.

point is simply that this case more clearly involves an interference with your freedom with respect to your mind than with your mind itself.)

The forgoing suggests that the aspect of the right to freedom of thought that protects against interventions on the mind is in some ways narrower and in other ways broader than the RAMI that I will defend in this book. However, this is an issue that remains to be settled. Though the moral right to freedom of thought has received greater attention in philosophical literature than the RAMI, my own view is that it also remains under-theorised and itself deserves a book-length philosophical treatment. This, however, is not that book; my focus will remain squarely on the RAMI.[18]

IV. Conclusion

In this chapter, I have offered a schematic, tentative, incomplete, and metaphysically ecumenical working account of the RAMI—one that parallels the account of the RABI offered in the previous chapter.

I first offered a preliminary account which consisted in three jointly sufficient conditions for infringement of the RAMI. I claimed that it is sufficient, for *A* to infringe *B*'s RAMI, that *A* is a well-situated moral agent, *A* significantly interferes with *B*'s mind, and *B* possesses the RAMI and has not waived or forfeited it in relation to the interference in question. I then deepened this account by offering a working account of mental interference. On this account, *A* interferes with *B*'s mind if *A*'s action *X* alters a mental state of *B*'s, and the mediating process—that is, the process within *B* via which *X* produces the alteration—is insensitive to the reasons that bear on the alteration. This account of the RAMI will be fleshed out in later chapters, and especially in Chapter 6. But for the moment, we have enough to work with. We can now turn to consider how the existence of this right might be justified.

[18] Although in offering a defence of the RAMI in the next chapter, I will consider various close relatives of it, some of which might be thought to be aspects of the right to freedom of thought.

4

An Intuitive, Abductive Defence of the Right Against Mental Interference*

In Chapter 1, I motivated the thought that we all possess a moral right against mental interference (RAMI); I suggested that the existence of such a right is *prima facie* plausible. In this chapter, I begin to develop a fuller defence of the existence of the right.

One of the considerations that I mentioned, in Chapter 1, as supporting the thought that we possess the RAMI, was that our possession of such a right is able to explain the wrongfulness of certain actions that seem, intuitively, to be wrongful. In this chapter, I begin to make the case for a stronger claim: it is not merely that we *can* account for the wrongfulness of certain actions by invoking the RAMI, we would be correct to do so.

I. The Case

I organise my discussion around a central case and several variants of it. Though we will not get all the way there in this chapter, my aim will be to establish that the best full explanation for the wrongfulness of the actions described in these cases will require appealing to the RAMI.

Here is the central case:

> *Brain Drug.* Following a recent period of marital tension, your friend, Lucien, has been feeling down; much of his mental life has acquired a more negative quality than it had previously. Activities that he previously enjoyed give him less pleasure, unpleasant activities seem more disagreeable, painful sensations are stronger, and so on. You think he might benefit from medical treatment, and you've tried convincing him to seek help, but he has proved resistant. So you take it upon yourself to put an end to his

* This chapter draws on some arguments that were initially published in T. Douglas (2025a).

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0004

> gloominess. You arrange to meet Lucien for coffee, and bring with you, in your backpack, a spray-bottle containing a newly developed drug. When sprayed on the skin, this drug is quickly absorbed into the blood and carried to the brain, where it attenuates feelings of gloom. It has no other mental or physical effects—that is, no effects not mediated by the diminished gloominess. A few minutes into your meeting with Lucien, you quickly lift the spray bottle from your bag and administer the drug by spraying it onto his hand, immediately thereafter explaining to Lucien what the drug is and why you administered it. And your intervention works: the drug causes Lucien to feel significantly less gloomy for the subsequent few days.

It seems to me that you wrong Lucien in *Brain Drug*—you breach a *pro tanto* duty owed to him.[1] But how do you wrong him? One suggestion might be that you wrong him by harming him. Perhaps Lucien will long be severely disturbed by your action. Perhaps your friendship will be destroyed, at great cost to Lucien. Perhaps Lucien's gloom would, had you allowed it to persist unabated, have motivated him to embark on a brilliant and successful career as a writer of Nordic noir. Or perhaps there are deep problems in Lucien's marriage to which his gloominess is an appropriate response, and which his gloom might ultimately have motivated him to resolve.

But let us suppose that your action has none of these effects, nor, indeed, any other harmful effects, and that it has at least some beneficial effects. Suppose, that is, that all of the consequences of your action, for Lucien, are non-harmful, and at least some are beneficial.[2] Finally, let us suppose that you intended and justifiably expected that the effects of your action would be overall beneficial for Lucien, and that none would be harmful.

Still, it remains plausible that you wrong Lucien. So again, we may wonder: how do you wrong him? If the effects of your action are overall beneficial and universally non-harmful for Lucien—both in justified expectation and in fact—then what is it that makes your action wrongful?

One answer—and the one I want to defend—holds that you wrong Lucien in *Brain Drug* (in part) by infringing his RAMI, as I specified that right in the previous chapter. But clearly other explanations are available. In this chapter,

[1] In holding that you wrong Lucien, I do not mean to imply that you perform an action that *is wrong*—where that would imply that it is, all things considered, impermissible. Wronging someone—or, as I will sometimes say, treating her wrongfully—can, I take it, be permissible, all things considered.

[2] This is consistent with your action *constituting*, rather than causing, a harm to Lucien. I am not here excluding that possibility.

I examine four possible alternative explanations. These hold that you wrong Lucien by, respectively, treating or affecting him in a way that he has not autonomously chosen (Section II), interfering with his body (Section III), reducing his mental autonomy (Section IV), and subjecting him to undue mental influence (Section V). I consider whether these explanations can, separately or cumulatively, fully and satisfyingly account for the wrongfulness of your action in *Brain Drug* and a number of related cases.[3] I allow that certain variants of the second and fourth explanations may be able to do so, but these variants imply the existence of the RAMI. I do not succeed in fully and satisfyingly explaining the wrongfulness *without* implying the existence of the RAMI.[4]

Before proceeding to my main argument, I need to offer a comment on a concept that will play only a limited role in my discussion, and whose limited role might otherwise seem an unjustified omission: the concept of manipulation. Your action in *Brain Drug* is, some would say, an instance of mental manipulation. Moreover, manipulation is often thought to be necessarily, invariably, or at least presumptively wrongful. So it might be thought that we can explain the wrongfulness of your action by adverting to its manipulativeness. However, I will not pursue any such explanation here. I eschew any such appeal to manipulation for three reasons. First, manipulation is a highly contested category. There is little agreement on the boundaries of the category of manipulation, on whether manipulation is always wrongful, or on why it is wrongful, when it is. Second, even if we could agree on the boundaries of the category, it is doubtful that the category would be sufficiently unified to have much explanatory power. I agree with J. S. Blumenthal-Barby when she writes that 'it is not the case that any of [the] practices traditionally labeled as "manipulation" are ipso facto morally wrong; nor is it even the case that any of these practices always has a single wrong-making feature' (2014, p. 122).[5]

[3] I use 'wrongfulness' to refer to the moral gravity or seriousness of a wronging, which I suspect will be determined by (i) the strength of the directed *pro tanto* duty whose breach the wronging consists in, and (ii) the magnitude of the breach of that duty.

[4] Though I will be looking for a full explanation of the wrongfulness of your action in *Brain Drug* and some variants thereof, the 'full' should not be taken to imply that the explanation has to bottom out in explanatorily fundamental considerations. Rather, the 'full' implies only that I want to explain all of the wrongfulness present in the cases I discuss. Thus, when I claim that the RAMI allows us to fully explain the wrong present in these cases, that should not be taken to imply that the RAMI is the end of the explanatory story. It may be that the RAMI can itself be explained by more fundamental considerations.

[5] In a similar vein, Sophie Gibert claims that '[t]here is no need to identify a common feature shared by all instances . . . of wrongful manipulation, for there is no unified, non-morally specifiable feature to be found' (2023, p. 336).

Third, and finally, there are some plausible and influential accounts of manipulation on which your intervention in *Brain Drug* is not—or at least, is not clearly—an instance of manipulation. For example, on Robert Noggle's (2020, 2025) account, manipulation consists in causing the manipulee to make some kind of mistake—for example, to have a false or irrational mental state—and it is not clear that you induce any mistake in Lucien.

I will, then, have very little to say about manipulation in this chapter, and indeed in the remainder of the book, though I will draw liberally on the very rich existing philosophical literature on manipulation.

II. Unchosen Treatment

It is natural to suppose that you wrong Lucien in a way that concerns his autonomy, and indeed the standard view about how nonconsensual medical interventions wrong their recipients is that they infringe or disrespect the recipient's autonomy. So let us start with the thought that you wrong Lucien by infringing some autonomy-right of his—some right he has to determine, control, or shape the course of his own life.

Various questions immediately arise. One of these concerns the *objects* of our right to autonomy: with respect to *which aspects* of our lives do we have a right to autonomy? A second question concerns the type of response that another person's right to autonomy calls for. Suppose, for example, that we have a right to autonomy with respect to our actions. Is this, for example, a right that others not *reduce* our autonomy of action? That they not *interfere* with it? That they not *disrespect* it? That they provide us with the wherewithal to exercise it? Yet another question concerns the distinction between 'autonomy-rights' and other rights. Which rights, exactly, count as autonomy rights? For instance, is the right against bodily interference an autonomy-right, since it allows a role for autonomous consent in determining the moral wrongfulness of an intervention on one's body, or does autonomy have to play some more central role in the content of a right for that label to be warranted?

I suspect that there are no easy answers to these questions. Thus, my strategy, in this chapter, will be simply to consider various different rights (i) that we plausibly possess, and (ii) that could plausibly explain the wrong you perpetrate against Lucien in *Brain Drug* and related cases. Two of the rights I consider—those discussed in this section and in Section IV—can clearly be aptly construed as 'rights to autonomy', and I will explicitly refer to

them as such. I do not take a stance on whether the other two rights—those discussed in Sections III and V—should also be construed in this way.

In this section, I consider the possibility that we have a right to autonomy with respect to how we are treated or affected by others. I take this to be a right that others allow us, within certain limits, to autonomously determine how they treat or affect us. Call this the right to autonomy of treatment. This right to autonomy of treatment might include a positive component (a right that others treat or affect us in certain ways that we have autonomously chosen to be treated or affected), and it very plausibly includes a negative component (a right that others not treat or affect us in certain ways that we have not autonomously chosen to be treated or affected). In what follows, I focus only on the negative aspect of the right, since that is the aspect at stake in *Brain Drug*, and it is also the less controversial aspect. I henceforth use the term 'right to autonomy of treatment' to refer only to this negative right.

I think it is plausible that we all possess something like a right to autonomy of treatment. And I think it is plausible that you infringe this right in *Brain Drug*. Lucien's right to autonomy of treatment includes a right not to be subjected, without his having autonomously chosen it, to your intervention in *Brain Drug*. Lucien has indeed not autonomously chosen your intervention, but you administer it anyway. Thus, you infringe his right.

However, it seems to me that this appeal to a right to autonomy of treatment can only be part of the story. To see why, note that we can often treat or affect a person in ways that they have not autonomously chosen, but without wronging them. If I befriend your best friend, or set up a shop next to yours, I affect your life, and perhaps very significantly, but I do not obviously wrong you, even if you had not autonomously chosen for me to do these things, and indeed even if you had autonomously chosen that I not do them. Similarly if a judge hands me a sentence, or my boss offers an unfavourable appraisal of my work, or a friend offers me unwelcome advice, these individuals 'treat me' in certain ways, but they do not clearly wrong me, even if I had never autonomously chosen to be treated in these ways, and indeed even if I had made it clear that I did not want to receive the sentence, the appraisal or the advice.[6]

[6] It might be thought that, in typical cases where one is given a criminal sentence, performance appraisal, or piece of advice, one will have made oneself liable to this by, for example, committing a crime, accepting a job, or embarking on a friendship. However, one can readily imagine cases in which this is not so. (Suppose the advice is given by an acquaintance whose friendship I have never encouraged.) And in such cases, it remains very doubtful that the treatment infringes any right.

For a case more closely analogous to *Brain Drug*, consider the following scenario:

> *Explicit Reasons*. All is as in *Brain Drug* except that, in this case, rather than spraying a drug on his skin, you seek to *persuade* Lucien out of his gloominess. Without prior warning, you present him with various reasons why he should be more cheerful; you point to aspects of his work that he enjoys, remind him that he has a loving family, emphasise the esteem in which you yourself hold him, and so on. Lucien finds your case compelling. He concludes that he indeed has many reasons to be more cheerful, and this somewhat diminishes his gloominess.

Your course of action in this case may be impolite and pushy. But does it wrong Lucien? I do not find it obvious that it does. Yet you treat Lucien in a way that he has not autonomously chosen to be treated.

It seems, then, that our right to autonomy of treatment—even when understood as a purely negative right—has its limits. It covers actions that treat or affect us in some ways, but not in others. What explains this? A plausible answer is that it is a parasitic right.[7] We possess various more fundamental rights against being treated or affected in certain ways. At least some of those rights are rights that we can waive by autonomously consenting to a form of treatment that would otherwise infringe the right. This is where the role for autonomy enters the scene.

When we possess a waivable right against treatment *x*, then whether we autonomously consent to treatment *x* will determine whether there is an infringement of that right. Thus, we possess a derivative right to autonomy of treatment with respect to treatment *x*. But the basic right here is the right against treatment *x*, not the autonomy-right.

The question then becomes: on what more fundamental right is Lucien's right to autonomy of treatment with respect to your brain stimulation intervention parasitic? Which deeper right of his do you infringe by nonconsensually modulating his gloominess as you do? In the subsequent three sections, I will consider three different answers to this question.

[7] Ripstein (2006, p. 221) makes an analogous point.

III. Interfering with the Body

As we saw in Chapter 1, it is commonly accepted that we all possess a right against bodily interference (RABI). I think it is somewhat plausible that you infringe this right in *Brain Drug*, and the working account of the RABI offered in Chapter 2 can accommodate that verdict. Spraying a substance onto Lucien's skin arguably counts as an instance of bodily interference via bodily touching, and pharmacologically altering the brain states on which Lucien's gloom supervenes arguably counts as bodily interference via bodily alteration. Moreover, though neither of these interferences seems large in magnitude, it is, I think, not implausible that their magnitude is sufficient to cross the threshold for 'significance'.[8]

In any case, let us grant, for the sake of argument, that *some* of the wrongfulness of your action in *Brain Drug* is explained by its infringing Lucien's RABI. The question I want to address now is whether we can explain *all* of the wrongfulness of your action in this way.[9] There is, I think, good reason to doubt that we can. The main problem is that the mental impact of your intervention just seems much more important, both morally and prudentially, than its bodily impact. It seems to matter quite a lot, both for Lucien and from a moral point of view, that you have changed Lucien's level of gloom. On the other hand, it seems not to matter so much, in itself, that you have caused some droplets to land on his skin and the chemical properties of some of his neurones to change a little. It would thus be strange if the latter impacts explained all, or even most, of the wrongfulness of your action. That would be rather like saying that the wrongness of branding someone with a hot poker is explained not by the pain it causes, but by the degree to which it heats up the skin or the activation of pain receptors and pain centres in the brain that it causes. It would be to confuse the physical causes and bases of the morally significant effect for the morally significant effect itself.

So it seems to me that the wrongfulness of your intervention in *Brain Drug* ought to be explained, at least in large part, by its *mental* impact. Of course, we could say—and a physicalist about the mind presumably would say—that the change you bring about to Lucien's mental states is itself also a bodily change, since mental states count as bodily states too. Thus, the most

[8] I will henceforth use the term 'brain' to refer to both the central and peripheral nervous systems.

[9] For consideration of a parallel issue in the law, see Blitz (2010, pp. 1103–5).

significant effects of your action on Lucien—the effects on his gloominess—can be brought within the purview of the RABI.

I have no objection to classing mental states as bodily states for the purposes of applying the RABI. However, to do so would, I think, be to allow the RAMI back into the picture. Recall that we were looking for an *alternative* to the RAMI-based explanation for the wrong present in *Brain Drug*. If we count mental states as bodily states for the purposes of applying the RABI, we are no longer considering an alternative to the RAMI-based explanation. Rather we are building the RAMI into the RABI. For note that if we take mental states to be bodily states for the purposes of applying the RABI, we could as well decompose the RABI, or at least the part of it protecting against bodily *alterations* (as opposed to touching), into two distinct subrights: the subright protecting against certain alterations to non-mental bodily states, and the subright protecting against certain alterations to mental states. It would then be the infringement of the latter subright that does most of the work in explaining the wrong present in *Brain Drug*. But this subright just is the RAMI.[10]

It seems to me, then, that we either understand mental states to be bodily states for the purposes of applying the RABI, in which case we have built the RAMI into the RABI, or we understand mental states as distinct from bodily states, in which case the RABI cannot plausibly account for all of the wrongfulness present in *Brain Drug*.

IV. Reducing Mental Autonomy

The appeal to the RABI may explain a part of the wrongfulness of your action in *Brain Drug*, but, unless we understand the RABI so as to include the RAMI, I do not think it can explain all of it. It cannot explain the wrongfulness that is attributable to the mental impact of your action on Lucien.

How can we explain *that* wrongfulness? One way to do so would be to invoke the RAMI. But there are other options available. In this section, I consider the possibility that we could explain it by invoking a right that others not reduce our mental autonomy, understood as our control over our minds.[11] The thought here is that you wrong Lucien in *Brain Drug* by

[10] I assume here that the significance condition contained in this subright of the RABI would pick out the same interferences as the significance condition contained in the RAMI.

[11] For examples of authors suggesting that autonomy consists (in part) in control, see Arpaly (2003, p. 118), Mele (2001, p. 9), Raz (1986, p. 269), and Taylor (2009, pp. 6, 7). For examples of

acting in a way that diminishes his control over (some aspect of) his mind.[12] This thought sits well with extant critiques of certain forms of mental influence. To give just two examples, Jan Christoph Bublitz and Reinhard Merkel (2009) suggest that the use of neurointerventions to alter personality traits might reduce one's autonomy of thought or action and be problematic for this reason, and so-called nudge interventions are sometimes criticised on the same basis.[13]

Two qualifications are called for here. First, the view that your action reduces Lucien's mental autonomy should be distinguished from the view, already discussed in Section II, that Lucien lacks autonomy with respect to how you treat him. As we saw in Section II, autonomy of treatment may play a role in the explanation of how you wrong Lucien, but it cannot be the full story, because we do not always have a right to choose how others treat or affect us. In any case, in this section, my focus is not on Lucien's level of control over your treatment of him. Rather, it is on his control over his own mind. And your treatment of him is just one of many factors that could affect his control over his own mind.

Second, though I will assume in this section that autonomy consists in control, I will remain neutral on whether being autonomous with respect to some aspect of one's mind requires only having the ability to control that aspect or requires also that this ability is in fact exercised. I remain neutral, we might say, on whether it is *having* control or *exerting* control that matters. Thus, for example, I do not take a stance on whether having autonomy with respect to some desire requires only that one has the capacity to control that desire, or that the desire was in fact produced or sustained by an exercise of this capacity.[14]

authors suggesting that it consists in part in self-governance or self-authorship, each of which is plausibly closely related to control, see Bratman (2005, p. 52, n. 1), Buss (2005, esp. pp. 195–235), Eidelson (2020, pp. 1637–38), Fischer (2012, p. 174), Kipper (2021, p. 2), Quong (2011, p. 45), and Raz (1986, p. 369).

[12] A variant of this view would hold that the wrongfulness of your action can be fully explained by the fact that it *should be expected* to reduce Lucien's control over some protected aspect of his body or mind. I avoid the need to consider this variant by assuming, throughout this section, that all of the effects of your intervention on Lucien are known to you in advance. Thus, what should be expected to occur is precisely what will actually occur.

[13] For example, Jens Kipper suggests that nudging by giving misleading information may 'restrict' or 'undermine' autonomy (2021, p. 9), and Neil Levy considers (and rejects) the view that nudges 'threaten' autonomy (2019, p. 281).

[14] I will also remain neutral on who or what, exactly, must have or exert this control. Suppose that Tana sometimes enters a psychotic state, and on one occasion, when in that state, decides to burn down his own house. Though in some sense, Tana's decision to burn down his house was controlled by an aspect of Tana, this does not suffice to make Tana autonomous with respect to

With these qualifications in hand, let us turn to consider whether your gloominess-reducing intervention in *Brain Drug* diminishes Lucien's control over some aspect of his mind.

We should note, to begin with, that there will plausibly be some aspects of Lucien's mind that your intervention affects, but without diminishing his control over that aspect. For example, though your intervention will presumably have some effects on Lucien's future decisions—perhaps his enhanced mood will lead Lucien to decide to take up a new hobby, or sign up for a dating app, or ask for a pay rise—it is not at all clear that his *control over* his future decisions as a whole, or indeed over any of his individual decisions, has been lessened. Indeed, it may be that your action has increased his control over many of his future decisions. Perhaps his gloominess was, by sapping his motivation, functioning as a constraint on what Lucien was able to decide. And perhaps, as a result, your intervention serves to enhance his decisional autonomy.

Still, it seems plausible that you have diminished Lucien's control with respect to *some* aspects of his mind. For example, we might think that Lucien has lost some of his control over how gloomy he feels.

Even here, there are difficulties, however. An initial problem becomes apparent once we recognise that Lucien's gloominess is temporally extended. Even if you have, in the short term, diminished his control over his gloominess, you may well, over a longer term, have increased it. Suppose it turns out that, in diminishing his gloominess, you contribute to Lucien entering a reflective and clear-thinking phase of his life, during which he finds that he is able to consciously foster and attenuate many of his affective states, including his feelings of gloom. The long-term effect of your intervention may thus be to enhance his control over how gloomy he feels.[15] If so, we could not explain the wrongfulness of your intervention by appealing to Lucien's net loss in this control, for he suffers no such loss.

A defender of the view that we can fully account for the wrong in *Brain Drug* by appealing to a reduction in Lucien's mental autonomy might

the decision. For an individual to have autonomy with respect to something, the right aspect of the individual—perhaps the 'true self'—must have or exert control. But there are various ways in which we might further characterise this proper locus of control, and I will remain neutral between these. When I refer to an individual as having or lacking control over something, these references should be understood as referring to whatever aspect of the individual must have control in order for the individual to be autonomous with respect to that thing.

[15] Similar points have been made by others. See, for example, Blumenthal-Barby (2012, p. 356), Blumenthal-Barby and Naik (2015), Quong (2011, p. 69), and Raz (1986, p. 419).

respond, at this point, by maintaining that you wrong Lucien in *Brain Drug* not by diminishing his *global* control over his gloominess, but by diminishing his *local* control over that gloominess—by which I mean, simply his control over his gloominess during some limited period of time. Perhaps, to take this response to its limit, she could maintain only that you have diminished Lucien's control over how gloomy he feels at precisely the moment that your interference first exerts its gloominess-mitigating effect. At that instant, Lucien's level of gloominess is, it might seem, beyond his control—or at least, less within his control than it would have been in the absence of your action.

However, it is possible to imagine versions of *Brain Drug* in which Lucien does not suffer even this reduction in control. We simply need to stipulate that Lucien's gloominess was already outside of his control and would have remained so even in the absence of your intervention. Perhaps, for example, the gloominess was triggered not by marital tension, but by a brain tumour caused by random genetic mutations. And perhaps this tumour is inducing the gloom through neural processes that Lucien was powerless to resist: the tumour causes him to experience some fixed level of gloom that he cannot alter; your drug reduces Lucien's gloom to a new, lower level. Call this variant of the case *Tumour*. Since, in *Tumour*, Lucien in any case already fully lacks control over how gloomy he feels, and will continue to lack such control, you have done nothing to diminish his control, at least, not relative to his prior state (call this the 'temporal baseline') nor to the state that otherwise would have obtained (the 'standard counterfactual baseline'). Nevertheless, I think that you wrong Lucien. It was not up to you to unilaterally modify his gloominess in this way.

How can we account for this wrong?[16] One way to do so would be to invoke a different baseline. We could say that you wrong Lucien by reducing his autonomy not relative to either the temporal baseline or the standard counterfactual baseline, but relative to some other, more idealised baseline, such as the situation that would have obtained had *nothing out of the ordinary* restricted his control over his gloom. However, there are other cases in which this response is not available.

[16] As Jonathan Pugh (2020, pp. 81–82) points out, *not reducing* a person's autonomy is consistent with wronging a person by failing to increase her autonomy. However, in *Tumour*, it seems that you wrong Lucien even if there is nothing else you could have done that would have left him with more control over his gloominess.

Consider a variant of *Brain Drug* which I will call *Dream Modulation*. In this variant of the case, Lucien's gloom afflicts him only during dreams. In these dreams, Lucien experiences himself to be feeling gloomy while lying on his back and watching dark clouds blow across the sky above him. These dreams occur frequently, and their gloominess is very consistent. It is not, for example, affected by what he has done during the preceding day, nor by the smells, sounds, or other environmental conditions that surround him as he sleeps. However, you then give Lucien a drug that greatly reduces the gloominess of these dreams; it replaces them with pleasant dreams in which he is lying on his back, feeling warm, and watching beautiful birds fly across the sky above him. Your intervention does not directly affect Lucien's wakeful mental life at all. Moreover, your action also has no indirect effects on his wakeful mental life, since you administer the drug while he is sleeping, he never remembers or otherwise experiences wakeful mental effects of his dreams, and he does not discover what you have been up to.

In this case, we cannot with justification say that you have reduced Lucien's control over his gloominess, even relative to an idealised baseline in which nothing out of the ordinary restricts this control; within-dream gloominess is simply not the sort of thing over which we, in normal circumstances, have control—at least not sufficient control for our dream states to count as autonomous. Moreover, it is, I think, hard to see any other plausible way of conceiving of the baseline such that there is a reduction in control in this case. Nevertheless, in subjecting Lucien to your dream-modulating intervention without his consent, you surely still wrong him, if only mildly; again, it was not up to you to unilaterally tinker with his dreams in this way. It seems that we cannot explain the wrong present in *Dream Modulation* by appealing to a loss in mental autonomy. Note also that it is doubtful that we can solve this problem by wheeling the RABI back in and appealing to the *bodily* interference involved in *Dream Modulation*, for in *Dream Modulation*, as in *Brain Drug*, it is plausible that at least part of the wrongfulness of your action is due to the mental impact of your action, not its non-mental bodily impact.[17]

[17] This discussion may have implications for our original case—*Brain Drug*—too. It is plausible that the unaccounted-for wrong present in *Dream Modulation* is a wrong that is also present in *Brain Drug*. Thus, it seems that we cannot *fully* account for the wrong present in *Brain Drug*; we cannot account for that part of the wrong that is shared with the unaccounted-for wrong in *Dream Modulation*.

V. Undue Mental Influence

In *Dream Modulation*, you do not reduce Lucien's control over any aspect of his mental life. Still, it seems clear that there is something problematic about the way in which you act on his mind. We might capture this thought by saying that you exert *undue mental influence* on his mind, where that term is, for the moment, simply a placeholder for whatever forms of mental influence wrong a person other than by causing harm, interfering with the person's body, or reducing the person's mental control.

In this section, I will distinguish various specifications of the right against undue mental influence and examine whether any of these can fully and satisfyingly account for the wrongfulness of your action in *Brain Drug*, its variants *Tumour* and *Dream Modulation*, and some further variants of these cases that I will introduce. I believe that we *can* fully and satisfyingly explain this wrongfulness if we understand the right against undue influence as, or as including, the RAMI—that is, as a right against (at least) significant mental interference. In the remainder of this section, however, my focus will be on determining whether there are other ways of going—ways of understanding the right against undue mental influence that can also account for the wrongfulness of your action in these cases, but without implying the existence of the RAMI.

In what follows, I consider four alternative formulations of the right against undue mental influence. In each case, I suggest that the formulation in question leaves the right against undue mental influence unable to fully and satisfyingly explain the wrongfulness present in either one or more of the cases I have already introduced, or in one or more new cases that I will introduce.

A. Undue Influence as Purely Physical Mental Influence

Consider first the possibility that undue mental influence should be understood as *purely physical* mental influence. I take purely physical mental influence to be mental influence in which the alteration to the target's mental states is produced via a process, within the influencee, that is not a mental process and contains no mental processes.[18]

[18] This view is inspired by a view expressed by Jan Christoph Bublitz and Reinhard Merkel. Bublitz and Merkel (2014, p. 69) propose that we possess a right against 'direct'

On this view, a right against undue mental influence could account for the wrongfulness of your intervention on Lucien in *Brain Drug* and its variants *Tumour* and *Dream Modulation*, since in those cases you alter Lucien's level of gloominess through a purely physical process—the process via which the drug you administer travels to, and chemically affects, the neural bases of Lucien's gloominess.

However, the right against purely physical mental influences fails to capture the wrongfulness of many intuitively wrongful mental influences whose wrongfulness can be accounted for by the RAMI. Consider a case of subliminal imagery. Suppose that you had caused Lucien to feel less gloomy not by administering a drug, but by confronting him with subliminal images of, say, smiling faces. Intuitively, this would be wrongful, but whereas the RAMI can plausibly account for the wrongfulness of this intervention, a right against purely physical mental influences cannot, since subliminal imagery exerts its effect via a process that includes mental processes, such as the (subliminal) processing of perceptual information.

Moreover, even in cases where a right against purely physical mental influences *could* explain the wrongfulness present, there is scope to question whether this explanation would be a satisfying one. This is because it is hard to see why it should make a difference to the moral appraisal of a mental influence whether it operates via purely physical or instead partly mental processes; it is not clear why it should matter whether you make Lucien feel less gloomy by electrically stimulating his brain, or through subliminal imagery, say. An appeal to the RAMI does not face this same problem, since it is not difficult to see why it should make a difference, to the moral appraisal of a mental influence, whether it operates via reasons-sensitive or reasons-insensitive processes. (The moral relevance of this distinction could, for example, be explained by invoking the thought that we ought to treat rational beings as rational beings and this requires influencing them via reasons-sensitive processes.)

interventions on the mind, which they characterise as 'stimuli reaching the brain by other routes than sensual perception'. Indirect interventions, by contrast, 'are perceived sensually (i.e. heard, seen, smelled, felt, even if not apprehended or reflected upon consciously) and pass through the mind of the person, being processed by a host of psychological mechanisms'.

B. Undue Mental Influence as Controlling Mental Influence

Another possibility to consider is that we should understand undue mental influences as *controlling* mental influences. Some object to mental influences that involve the exercise of control by the influencer over the influencee. For example, some object to the use of highly targeted online advertising on the basis that it makes the advertising viewers 'puppets' (Véliz 2020, p. 71) or 'targets for remote control' (Zuboff 2020). More generally, it has been argued that the degree of control exerted by an influencer over an influencee is, at least often, relevant to the moral assessment of that influence.[19] Thus, for example, coercion and (some forms of) manipulation are sometimes thought to be morally problematic in part because they amount to control of one person by another.[20]

It is not always clear what authors have in mind when they characterise mental (or behavioural) influences as 'controlling'. In some cases, the worry seems to be that the influence is difficult or impossible for the influencee to resist.[21] In these cases, the concern is, I think, primarily about the influencee's lack of control with respect to one or more of the mental alterations that the influence produces, not the influencer's possession of control. (For reasons to which I will return, it is possible that the influencee could lack control over the alteration, say because the influence is difficult to resist, without the influencer having control over it.) I will consider a version of this worry in the next subsection. In this subsection, I focus on the worry that the influencer exercises too much control. More specifically, I consider the possibility that

[19] See, for example, Saghai (2013, p. 489), who writes that 'an influence preserves freedom of choice if and only if it preserves the choice-set and is fully or substantially noncontrolling'. See also Faden et al. (1986, esp. pp. 238, 256–58), who argue that the absence of controlling influences is a necessary condition for autonomous action.

[20] See, for example, Faden et al. (1986, ch. 10).

[21] This appears to be the case for both Saghai (2013) and Faden et al. (1986), who frequently contrast controlling influences with easily resistible ones and arguably suggest that the resistibility of an influence is the only or at least most important determinant of whether it counts as a controlling influence. Saghai claims that 'A's influence to get B to ϕ is substantially noncontrolling when B could easily not ϕ if she did not want to ϕ' (p. 488) and that '[r]esistibility is a criterion for testing the degree to which an influence is controlling' (p. 488). Faden and collaborators write that '[i]n analyzing the condition of noncontrol, we use influence and resistance to influence as basic concepts in the analysis. Control is exerted through influences in our analysis, but not all influences are controlling. Many influences are resistible . . .' (p. 256); that '[p]erson A's action controls an action X of person B if A gets B to do X through irresistible influences that would work even if B, left to his or her own ends, in no way wanted to do X' (p. 258); and that '[i]t can be a tedious and thorny chore to decide whether psychological manipulations like low-balling are compatible with substantial noncontrol in specific cases . . . the relevant criterion is whether the manipulation is easily resistible by the manipulatee' (p. 367).

we each possess a right against those mental influences in which the influencer exercises excessive control over the influencee. I henceforth refer to such influences as 'controlling mental influences'.

Our possession of a right against controlling mental influences may again allow us to explain the wrongfulness that seems to be present in *Brain Drug*, *Tumour* and *Dream Modulation*. In those cases, you perhaps exercise a high degree of control over Lucien's level of gloominess. However, it fails to capture the wrongfulness of some other interventions that seem intuitively wrongful, and whose wrongfulness the RAMI *can* capture. Consider a variant of the *Dream Modulation* case. Suppose the pharmacological intervention that you use to mitigate the gloominess of Lucien's dreams is rather hit and miss. This drug sometimes reduces a person's gloominess, but often fails, and when it does diminish gloominess, it sometimes does so to a high degree, and sometimes only to a low degree. It seems to me that, were you to employ this blunt and fickle pharmacological intervention on Lucien, and to succeed in mitigating his gloominess, you would wrong him. We can account for this wrong by positing the RAMI, but we cannot account for it by positing a right against controlling mental influence, for, in this case, your intervention is not sufficiently reliable to qualify as an exercise of control.[22]

C. Undue Mental Influence as Hard-to-Resist Mental Influence

Consider a third possibility: undue mental influence should be understood as *hard-to-resist* mental influence.

The difficulty with which a mental influence can be resisted is plausibly relevant to the moral assessment of that influence. This is suggested by the literature on nudging, in which it is frequently held that nudges are morally innocuous in part because they are—perhaps by definition—easy to resist.[23] Difficulty-of-resistance is also sometimes invoked to explain the wrongfulness of and/or threat to autonomy posed by coercion and (certain forms of) manipulation.[24]

[22] For the view that control over something requires the ability to reliably determine the state of that thing, see De Marco and Douglas (2024).

[23] Thaler and Sunstein (2008), the most prominent defenders of nudges, claim that '(t)o count as a mere nudge, an intervention must be easy and cheap to avoid' (p. 6). For further development of this view, see Saghai (2013).

[24] See, for example, Faden et al. (1986, Chs. 7 and 10).

In what follows, I will assume that there are two descriptively different, but morally equivalent, ways of resisting an influence. Suppose someone performs action *M* and that *M* could, under certain circumstances, cause some mental alteration *N* in you. One way of resisting this influence—or, more accurately, this *potential* influence[25]—is to avoid exposure to action *M*.[26] (This would perhaps more naturally be described as *avoiding* than *resisting* the influence, but I treat it as a subtype of resisting.) Thus, for example, if you simply refuse to use a television that has been set to display subliminal images, you have resisted that influence. Another way to resist this influence would be to avoid undergoing alteration *N* but to do so without avoiding exposure to action *M*. For example, you could allow yourself to be exposed to subliminal images but use some psychological technique to block (some of) their possible mental effects.

An influence could be easy (by which I here mean simply 'not hard') to resist because it is easy to resist in one or other of these ways—it is not necessary for it to be easy to resist in both ways. Thus, I take it that for a mental influence to be 'undue', on the current proposal, it would need to be hard-to-resist in both ways.

A right against undue mental influence, understood as a right against hard-to-resist mental influences, may be able to account for the wrongfulness present in *Brain Drug*, *Tumour*, and *Dream Modulation*. However, the appeal to this right has similar downsides to the appeal to a right against controlling mental influences: it seems that many of the influences that I have posited as infringements of the RAMI would remain wrongful even if they were easy to resist.

Consider again *Dream Modulation*, in which you mitigate the gloominess of Lucien's dreams by administering a drug. Suppose that Lucien knows that you plan to implement this intervention on a given night, and that he could easily thwart this plan simply by locking his bedroom door before he sleeps on that night. However, suppose that he doesn't get around to locking his door, and you implement the intervention as planned. Nevertheless, Lucien made it clear that he did not consent to the intervention; he did not agree to your performing it. It seems that in this case you wrong Lucien, though

[25] In the remainder of this subsection, for ease of exposition, I refer to actions that would have counted as mental influences had they not been successfully resisted as 'influences', though 'potential influence' would be more accurate.

[26] I leave it open what it takes to be 'exposed' to an action, but I take it to be plausible that being exposed to an action could consist in witnessing it, or one of its external-to-the-nudgee effects.

he could easily have resisted the influence. The RAMI can account for the wrongfulness of this intervention, but a right against hard-to-resist mental influences cannot.

Alternatively consider a case in which you seek to boost Lucien's mood while he is awake by setting the television in your shared living room to display subliminal images of cute animals on one particular evening. Suppose that Lucien is aware of your plan to implement this intervention and explicitly refuses to consent to it. But suppose that he also knows that he could easily nullify the effect of the intervention simply by thinking briefly of how dreary the weather has been of late before he turns on the television that evening. He does not do this, however. Thus, your intervention indeed boosts Lucien's mood.

Again, it seems that you wrong Lucien in this case, but again, it seems that we cannot account for this wrong by invoking a right against hard-to-resist mental influences. This intervention is easy to resist. On the other hand, we plausibly *can* account for the wrongfulness of the intervention by invoking the RAMI.[27]

D. Undue Mental Influence as Nontransparent Mental Influence

The final possibility that I want to consider is that undue mental influence should be understood as *nontransparent* mental influence.

It has often been suggested, especially in the literature on nudging, that the transparency of an influence can be relevant to its wrongfulness.[28] Transparency means different things to different protagonists in this literature.[29] For example, for some, an influence must be announced in advance to count as transparent, whereas for others, what matters is that the influence can—in some sense or other—be detected by the influencee.

[27] In Chapters 6 and 7, I will examine whether interventions similar to this one indeed infringe the RAMI, and will suggest that they do.

[28] See, for example, Barton and Grüne-Yanoff (2015, p. 347), Blumenthal-Barby (2021, p. 119), Hausman and Welch (2010, p. 135), Ivanković and Engelen (2019, p. 59), Mills (2015, p. 499ff.), Sunstein (2014, p. 148), and Thaler and Sunstein (2008, p. 247).

[29] For discussion, see De Marco and Douglas (2023) and Ivanković and Engelen (2019).

A right against nontransparent mental influence will have the best prospects of accounting for the wrong present in *Brain Drug* and its variants if we adopt a demanding conception of transparency, and thus an inclusive account of nontransparency. So I will begin by supposing that transparency requires all of the following: that the influencing action and the likelihood that it will produce the resulting mental alteration are announced in advance, that both the influencing act and the resulting mental alteration can be easily detected by the influencee when they occur, and that it can easily be established by the influencee that the mental alteration is the result of the influencing action. Thus, an influence will count as nontransparent when (i) the influencing action or the likelihood that it will produce the mental alteration is not announced in advance, (ii) the influencing action or the resulting mental alteration cannot easily be detected by the influencee, or (iii) the influencee cannot easily establish that the alteration is the result of the influencing action.

If transparency is understood thus, a right against nontransparent mental influence could plausibly explain the wrong present in *Brain Drug*, *Tumour*, and *Dream Modulation*. As I initially specified these cases, you did not announce your plans in advance. Moreover, in one of these cases, *Dream Modulation*, you administered the drug while Lucien was sleeping (so it would presumably be difficult for him to detect your influencing action) and the drug's mental effects are present only during his dreams (so there is also a sense in which it is not easy for Lucien to detect the effects of your intervention on his level of gloom—*wakeful* Lucien has no access to these effects). In the other two cases—*Brain Drug* and *Tumour*—Lucien *can* detect both the influencing action (the administration of the drug) and the reduction in gloominess that it produces, but it is doubtful that he could easily establish that the reduction in gloominess is due to the influencing action. And in any case, the absence of prior warning is already enough to ensure that these interventions are not transparent, on my demanding conception of transparency. Thus, in all three of these cases, we could seemingly explain the wrongfulness of your action by appealing to a right against nontransparent mental influence.

Problems arise in relation to other variants of these cases, however. Consider a case—call it *Transparent Brain Drug*—in which you tell Lucien in advance that you are going to spray a gloom-attenuating drug onto his skin, and you then administer the drug in plain sight; you spray the drug onto Lucien's skin while he is awake and can see perfectly well what you are doing.

Suppose also that he can detect its effects on his gloominess, since in this version of the case, the drug mitigates Lucien's gloom during periods while he is wakeful, and that the drug-induced feelings of gloom have an unusual character, which Lucien knows shows that they were caused by the drug. Lucien makes it clear that he does not consent to your administering the drug, but you do so anyway.

In this version of the case, your influence on Lucien is transparent; it is disclosed in advance, Lucien can easily detect both your influencing action and the levels of gloom that it produces, and he can easily establish that the latter are the result of the former. So it will no longer infringe the right against nontransparent mental influences. Yet your intervention nevertheless seems wrongful. A right against nontransparent mental influences cannot, it seems, explain this. The RAMI, however, could.

At this point, it might be objected that my revised, more transparent case is still not sufficiently transparent. Your intervention in this case is transparent as I defined transparency above. But it might seem that there are other important ways in which it is not transparent. For example, Lucien cannot easily detect the motives behind your actions. And nor can he easily detect the processes via which the drug exerts its effects on his dreams. So perhaps even in the revised version of the case, you do still infringe Lucien's right against nontransparent mental influences.

However, if transparency is understood in this even more demanding way—as requiring motive transparency (that the influencee can easily detect the influencer's motives) and/or mechanistic transparency (that the influencee can easily detect the mechanisms via which the influencing act produces the mental alteration)—then the right against nontransparent mental influences will be even more inclusive, and indeed will cover an implausibly broad range of influences. For we lack motive and/or mechanistic transparency with respect to the vast majority of the mental influences to which we are subject. (Of course, it might be thought that the RAMI faces a similar difficulty; it too implicates too many mental influences. I will respond to this worry in Chapter 8.)

Thus, it seems to me that, if we are to avoid an overly broad understanding of the right against nontransparent influences, we will have to eschew any conception of transparency that requires motive and/or mechanistic transparency. But then we are back with our original problem: the right will be unable to account for the wrongfulness present in more transparent variants of *Brain Drug*, such as *Transparent Brain Drug*.

VI. Combining the Explanations

We have now considered four candidate explanations for the wrongfulness of your actions in *Brain Drug* and several variants of this case. These appealed respectively to the claims that you treat Lucien in a way that he has not autonomously chosen, interfere with his body, reduce his mental autonomy, and/or unduly mentally influence him. I have allowed that two of these explanations—the second and the fourth—might be able to fully and satisfyingly explain this wrongfulness if they are understood in a way that implies the RAMI. But I have not been able to identify any non-RAMI-implying variant of these explanations that can achieve the same result.

One possibility that I have not yet fully considered, however, is that some combination of the explanations that we have considered might succeed where these explanations, taken separately, have failed.

I do not think this approach will work. To see why, consider again *Transparent Brain Drug*, introduced in the previous section. In this variant of *Brain Drug*, you tell Lucien in advance that you are going to spray a gloom-attenuating drug onto his skin, and you then administer the dream-modulating drug in plain sight; you spray the drug onto Lucien's skin while he is awake and can see perfectly well what you are doing. He can detect its effects on his gloominess, since in this version of the case, the drug mitigates Lucien's gloom during periods while he is wakeful, and the drug-induced feelings of gloom have an unusual character, which Lucien knows shows that they were caused by the drug.

Suppose, now, that we further alter this case by adding the following modifications and provisos:

- Your intervention employs not a drug, but a pattern of flashing lights which exert their effects via perceptual processes.
- Though he is wakeful when the flashing lights intervention is administered, Lucien's level of gloom prior to your intervention is, and would in any case have remained, beyond his control, and not due to some unusual situation such as the presence of a brain tumour.[30] (We could suppose that Lucien has tried, without success, all of the strategies that might plausibly diminish his gloom.)

[30] This stipulation is intended to rule out the possibility that the wrongfulness of your action might be explained by the resulting reduction in mental autonomy.

- Lucien could easily have resisted your intervention, for example, by walking away from you after you announced your plans to use the procedure.
- Your flashing lights procedure is not reliable; it sometimes diminishes a person's level of gloom, as it does with Lucien, but it often unpredictably fails to do so.

Call the case that results from the addition of these modifications and provisos *The Heavily Modified Case*. I find it plausible that you at least mildly wrong Lucien in *The Heavily Modified Case*, yet you do not reduce his mental autonomy, and nor do you exert a purely physical, controlling, hard-to-resist or nontransparent mental influence (unless transparency is, implausibly, understood to require motive or mechanistic transparency). You do perhaps wrong him by interfering with his body, but it seems implausible to think that the non-mental bodily impact of the intervention fully accounts for its wrongfulness; its *mental* impact seems more normatively significant. Thus, I think that an appeal to the RABI can *fully* account for the wrong in this case only if the RABI is defined so as to include the RAMI, such that mental interferences count as bodily interferences.

I contend, then, that, even if we combine the various non-RAMI-implying explanations considered in this chapter, we will not be able to fully and satisfyingly account for the wrong present in *The Heavily Modified Case*.

VII. Conclusion

In this chapter, I have considered whether we can give a full and satisfying account of the wrongfulness of your action in *Brain Drug*, and several variants thereof, without invoking the RAMI. More specifically, I have considered whether we can fully and satisfyingly account for this wrongfulness by appealing to the claims that, in these cases, you treat Lucien in a way that he has not autonomously chosen, interfere with his body, reduce his mental autonomy, and/or unduly mentally influence him.

The first explanation provides an explanation that is not fully satisfying, since it does not distinguish your intervention on Lucien from other ways of affecting or treating him that, though unchosen by him, appear not to wrong him. The second explanation—the appeal to the RABI—may account for some of the wrongfulness present in *Brain Drug*, but it cannot

account for the wrongfulness that is attributable to the mental, as opposed to the non-mental bodily, impact of the intervention. At least, it cannot account for this wrongfulness unless we class mental states as bodily states, in which case the RABI will subsume the RAMI. The third explanation—the appeal to reduced mental autonomy—may allow us to explain some of the residual, as-yet-unexplained wrongfulness present in *Brain Drug*. But it cannot account for the wrongfulness present in *Dream Modulation*, since in that case Lucien experiences no loss in mental autonomy. The fourth explanation—which appeals to a right against undue mental influence—may allow us to fully and satisfyingly account for the wrongfulness present in *Dream Modulation*, and indeed in the other variants of *Brain Drug* that I considered, provided that the right against undue influence is understood as, or so as to include, the RAMI. However, all of the alternative formulations of the right that I considered failed to fully and satisfyingly account for the wrongfulness present in either these cases, or others that I introduced. Importantly, moreover, combining the various non-RAMI-implying explanations considered in this chapter did not resolve the problems faced by them separately; there are cases in which no combination of these explanations can fully account for the wrongfulness that is, intuitively, present.

In summary, then, I have not been able to fully and satisfyingly explain the wrongs present in the range of cases discussed in this chapter *without* positing a right that is or includes the RAMI. This, I think, should push us towards positing the RAMI, since that right seemingly *can* fully and satisfyingly account for the wrongfulness that is intuitively present in all of the cases I have introduced so far.

The argument offered in this chapter is, however, importantly limited. For all I have argued, it is possible that my proposed explanation for the wrong in *Brain Drug* and the other cases considered—the appeal to the RAMI—itself runs into problems. The most obvious worry, I think, is that my explanation may seem to show too much. It might be thought to imply, for instance, that some rather mundane forms of mental influence are wrongful. Consider, for example, influences such as wearing perfume on a date or playing mood music in one's restaurant. These influences seem intuitively rather innocuous, yet there is at least a case to be made that they infringe the RAMI, as I have specified it. I will address this type of worry in Chapter 8.

Pending this discussion in Chapter 8, I can, for the moment, offer only a rather limited conclusion. My claims, at this point, are only that (i) the

alternative explanations considered in this chapter cannot, unless understood so as to imply the existence of the RAMI, fully and satisfyingly account for the wrongfulness of your action in *Brain Drug* and its variants, and (ii) the appeal to the RAMI remains on the table as an explanation of the as-yet-unaccounted-for wrongfulness.

5

From Self-Ownership to the Right Against Mental Interference

In this chapter, I sketch out a theoretical defence of the right against mental interference (RAMI). The defence attempts to derive the RAMI from our ownership of ourselves, so I call it the 'argument from self-ownership'.

The discussion proceeds as follows. In Section I, I set out the argument from self-ownership and motivate—without fully defending—each of its three starting premises. I then, in Sections II–IV, consider and respond to three objections to this argument. I argue that none of these objections undermines the argument.

I. The Argument from Self-Ownership

The argument from self-ownership runs as follows:

[1] We own ourselves (the self-ownership thesis).
[2] If we own ourselves, then we own our minds.
[3] We own our minds (from [1] and [2]).
[4] If we own our minds, then we possess a right against interference with our minds.
[5] We possess a right against interference with our minds (from [3] and [4]).

In this section, I motivate the three premises of this argument: claims [1], [2], and [4]. I try to show that each is plausible and at least implicitly widely endorsed. Moreover, I try to do this without committing to any particular conception of self-ownership. I want to show that these premises should be and are accepted by proponents of a range of different accounts of self-ownership.

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0005

A. The Self-Ownership Thesis

Claim [1]—the self-ownership thesis—asserts that we own ourselves. What does it take to own oneself? As just indicated, I want to be open-minded about this. In what follows, I will sometimes need to make substantive claims about what self-ownership does or does not imply. However, in making such claims, I will try to accommodate as many of the dominant conceptions of self-ownership as I can. More specifically, I will avoid committing to a number of theses that that are sometimes, but not always, associated with the term 'self-ownership'. Some believe that self-ownership is absolute (in the sense that infringements of it are never justified) and ungraded (in the sense that all infringements of it are equally serious),[1] but I do not assume (or, as it happens, believe) that self-ownership has either of these features. Many also believe that self-ownership implies or includes a set of very rigid constraints on the permissible redistribution of resources between individuals,[2] but again, I do not assume (or, as it happens, believe) that self-ownership includes or implies these constraints.[3]

The self-ownership thesis has a long pedigree in moral and political philosophy. It has often been invoked in explanations of how property rights over external objects come into existence. For instance, self-ownership plays a central role in the standard Lockean account of initial property acquisition:

> The *labour* of his body, and the *work* of his hands, we may say, are properly his. Whatsoever then he removes out of the state that nature hath provided, and left it in, he hath mixed his *labour* with and joined to it something that is his own, and thereby makes it his property. (Locke 1982, Ch. V, §27 (emphasis in original))

The thought here is that we come to own external objects that we produce with our own labour because, in so producing them, we mix something we own (ourselves, which, on Locke's view, includes our labour) with something

[1] For a discussion of self-ownership that endorses both its absoluteness and its non-gradedness, see Flanigan (2019).

[2] I take this to be the dominant position in recent scholarship on self-ownership. For classic statements of it, see Nozick (1974) and G. A. Cohen (1995).

[3] My own view is that self-ownership is consistent with the permissibility of highly redistributive policies. This is in part because I think that rights of self-ownership are not absolute, and in part because I think their implications for the acquisition and transfer of external property are subject to strong fairness-based limits.

we do not own (natural resources), and the ownership we enjoy over our selves is inherited by the product of this mixing.

Self-ownership is also widely asserted as the basis for bodily rights, including the right against bodily interference. This right is often said to derive from, or to be an element of, self-ownership.[4]

So the claim that we own ourselves—the self-ownership thesis—is often endorsed. Unfortunately, what is less clear is *on what basis* it is endorsed. In philosophical discussions, the thesis far more commonly serves as a premise than as a conclusion. Nevertheless, it is, I think, possible to identify several plausible justifications for the self-ownership thesis.

Some of these justifications advert to some value that self-ownership, or recognition of it, realises or helps to realise. For instance, one way to justify the self-ownership thesis is to argue that recognising self-ownership is instrumental to protecting an individual's autonomy or well-being. Alternatively, the self-ownership thesis might be defended on the basis that the existence or recognition of a property right over the self contributes constitutively, not instrumentally, to some value. This view has a strong pedigree in non-consequentialist moral philosophy. In this literature, certain moral rights are often said to be (partly) constitutive of a particular moral status (often that of *inviolability*) or form of authority, the existence or recognition of which is taken to possess a special kind of noninstrumental value, either for the right-holder or impersonally.[5] Property rights over the self may be among those rights that are partially constitutive of this valuable moral status or form of authority.[6]

Besides these appeals to more fundamental values, there are also other, less lofty grounds on which we might endorse the self-ownership thesis. For example, we might endorse it on the basis of abductive considerations; we might posit self-ownership because it helps to explain things that are otherwise difficult to explain. For example, if we accept something like Locke's 'mixing'-based account of initial property acquisition, self-ownership might

[4] See, for example, Flanigan (2019, p. 10). Most often, it is simply assumed that self-ownership implies the RABI, but for a rare explicit argument for this claim, see Russell (2010).

[5] See, for example, Kamm (1992, pp. 382–86, 2004, pp. 492–95), Nagel (2004, pp. 33, 36–40), and Owens (2019).

[6] Indeed, it has been argued that rights over *external* property might also be defended in this way. For example, Owens (2019, p. 271) argues that property rights in general can be defended in part by reference to the constitutive contribution they make to our authority over that which we own: 'Individual ownership puts people in authority over the things around them, it is good for individuals to have authority of that sort, and a system of personal property can be justified, at least in part, by the fact that it serves that interest'.

help to explain how people initially came to acquire external property. Yet another possibility is that we might endorse the self-ownership thesis merely on the basis of its direct intuitive appeal—because it seems self-evident that we own ourselves.

There is no need for us to decide between these alternative justifications for the self-ownership thesis here. And it may be that two or more of them can be combined. All I wish to claim at this point is that there are multiple *prima facie* plausible grounds for endorsing the self-ownership thesis.

B. From Self-Ownership to Mind Ownership

Consider next the second premise in the argument from self-ownership: claim [2], the claim that, if we own ourselves, then we own our minds. This claim has not, to my knowledge, received any detailed philosophical attention. However, it is sometimes asserted. Indeed, self-ownership is sometimes defined as simply the conjunction of body and mind ownership. John Stuart Mill characterises personal sovereignty—arguably a form of self-ownership—thus: 'over himself, over his own body and mind, the individual is sovereign' (1909, p. 19). Kasper Lippert-Rasmussen likewise takes self-ownership to consist in a person's 'moral ownership of himself or herself, that is, his or her body and mind' (2018, p. 142), while Michael Otsuka understands it as '[a] very stringent right of control over and use of one's mind and body' (1998, p. 69). However, and perhaps surprisingly, when self-ownership is not *defined* so as to include mind ownership, its implications for the mind are seldom made explicit. Discussions of the implications of self-ownership typically focus on its implications not for the mind, but for the body, for one's labour, or for the acquisition of external property.[7]

Nevertheless, I believe that self-ownership will plausibly include mind ownership even when it is not defined so as to include it.[8] At least, I think this

[7] Moreover, explicit definitions of self-ownership often refer to the body and/or labour, but not to the mind. John Christman, for example, defines self-ownership as a property right over one's 'body, skills, and labor' (1991, p. 28; see also 1994, p. 147).

[8] Jan Christoph Bublitz and Reinhard Merkel (2014, p. 62) make a similar point: 'In the wake of Locke, libertarians believe that persons have property rights in their body; persons literally own (the physical part of) themselves. Ownership discussions focus on the relation of persons to their bodies, their liberty (e.g. vis-a-vis slavery) and the fruits of their labor. But what is even more constitutive of a subject than her body is her mind. So, whoever grants self-ownership of persons over their bodies has a compelling reason to concede self-ownership over minds.'

will be so *if* self-ownership is understood—as it typically is—so as to cover the whole body.[9]

To make good on this thought, I need to distinguish two different approaches to delineating the boundaries of the self. On one approach, the self referred to in 'self-ownership' is an entity whose boundaries can be determined prior to any normative discussion about what rights we have. On this way of thinking about things, we can first determine the boundaries of the self, and then determine whether we own the self.

On a second approach, 'self', as it figures in 'self-ownership', refers not to some entity that can be established on independent grounds, and which then fixes the scope of our putative self-ownership, but to a normative construct that falls out of the arguments for and against self-ownership. On this way of thinking about things, we first establish what property rights we have over our bodies and minds, or parts thereof, and we then use the term 'self' to pick out the entity covered by those rights.

I doubt that either approach will lead us to a conception of self-ownership on which it covers the whole body, but not the whole mind. Take the first approach first. When the self (or as I will here take to be equivalent, the person) is delineated independently of self-ownership, it is normally taken to consist in either (i) the whole human organism (as in animalistic accounts of personal identity, see, e.g. Olson 2003), (ii) the mind and its physical basis (as in 'embodied mind' conceptions of personal identity, see, e.g. McMahan 2002), (iii) the mind alone (as in psychological conceptions of personal identity, see, e.g. Locke 1849), or (iv) some part of the mind, such as the 'true self' or the 'rational agent' (see, e.g. Korsgaard 2009).[10] Notably, only the first of these views would allow for the self to include the whole body, and that view—as it is normally understood—would also include the whole mind within the self; proponents of animalist accounts of the person typically allow that the

[9] Some authors do consider the possibility that our property-like rights over our bodies may not cover our whole bodies. See, for example, Kasper Lippert-Rasmussen (2008, esp. pp. 89, 91) and Japa Pallikkathayil (2017, p. 52). Other authors have noted that the assumption that self-ownership implies ownership of the whole body is often too quickly assumed. See, for example, Peter Vallentyne (1998, pp. 613–14, n. 2): 'Cohen assumes, as is standard, that self-ownership, as applied to moral agents, includes ownership of one's entire body, but this does not follow from the thesis of self-ownership alone. It depends on facts about personal identity. If, for example, agents are purely mental beings that can, but need not, voluntarily occupy and leave bodies, then bodies (in their initial state) are simply (external) natural resources, just like tracts of land.'

[10] I assume here for the sake of argument that accounts of the person are also accounts of the self. This is the charitable assumption to make since it allows us to include a broad range of views in our search for a view of the self on which it includes the whole body but not the whole mind.

animals that we are can think—we are animals with minds (see, e.g. Olson 2003, §4).

Consider now the second approach, on which the boundaries of the self are dependent on normative considerations regarding the scope of our rights. I think that this approach will also likely lead us to the result that, if self-ownership covers our whole bodies, then it also covers our whole minds. This is because none of the grounds for endorsing self-ownership reviewed in the previous section privileges the body over the mind; none suggests that the body has any stronger claim to being included within the protection of this right than does the mind. For example, there is, it seems to me, no clear reason to think that a moral property right over our bodies would promote our well-being, promote our autonomy, or contribute constitutively to our inviolability to any greater degree than would a moral property right over our minds.

C. From Mind Ownership to a Right Against Mental Interference

Consider next the final premise in the argument from self-ownership: claim [4]—the claim that, if we own our minds, then we possess a right against interference with our minds.

This premise can be defended by appealing to the more general claim: if we own *x*, then we possess a right against interference with *x*. As with claims [1] and [2], this claim is not typically explicitly defended; however, it is commonly asserted or assumed. Property rights are normally understood to be bundles of subrights—rights of control, rights of use, rights to transfer, and so on—and one of those subrights is almost always taken to be a right against interference with that which is owned.[11] Indeed, some have taken this subright to be the essence of property.[12] That property includes this subright is also suggested by metaphors commonly used to describe property, such

[11] This subright, 'the claim that others should not without permission interfere', appears in what has arguably become the standard analysis of property—that due to A. M. Honoré (1961, p. 371).

[12] F. S. Cohen (1954, p. 371) claims that '[p]rivate property may or may not involve a right to use something oneself. It may or may not involve a right to sell, but whatever else it involves, it must at least involve a right to exclude others from doing something'. Taylor (2004, p. 67) asserts a stronger claim: 'the incidents (claim-rights) that constitute self-ownership impose perfect duties of physical noninterference upon others—that is all'.

as the metaphor of property rights providing a 'fence' around that which is owned.[13]

As well as being commonly asserted or adopted, the claim that owning *x* entails a right against interference with *x* is also, I think, a plausible one, for the simple reason that it is difficult to come up with counterexamples to it. Perhaps I suffer from a lack of imagination, but I can think of no instances in which a person clearly has a moral property right over something *without* having a moral right against interference with it. Of course, there are cases in which the right against interference entailed by a property right is ineffective, in the sense that it can be permissibly infringed. Suppose that I own a vial of some extremely dangerous radioactive substance. But suppose that some other person—an eminent chemist, say—is the only person able to stop this substance exploding, and that, to do this, she must constantly tinker with the substance, adjusting its temperature. It seems clear that this chemist could interfere with the substance without impermissibly infringing my property right over it. However, it also seems to me that I do still have a defeasible right against this interference.

D. Taking Stock

Recall that the argument from self-ownership runs as follows:

[1] We own ourselves (the self-ownership thesis).
[2] If we own ourselves, then we own our minds.
[3] We own our minds (from [1] and [2]).
[4] If we own our minds, then we possess a right against interference with our minds.
[5] We possess a right against interference with our minds (from [3] and [4]).

I have suggested that claims [1], [2], and [4] would each garner considerable (implicit or explicit) support and that each has some initial plausibility. Moreover, claim [3] follows straightforwardly from [1] and [2] by *modus ponens*, as does [5] from [3] and [4]. Do we, then, have a compelling

[13] For an example of this metaphor being drawn in relation to self-ownership, see Christman (1994, p. 150), who characterises one conception of self-ownership as providing 'an argument that a kind of fence be built around people to keep other people out'.

argument for the RAMI? Not yet. The argument from self-ownership is potentially vulnerable to at least three objections—or, more accurately, groups of objections—that I will outline and assess in the next three sections.

II. Against Self-Ownership

An initial group of objections target claim [1], the claim that we own ourselves. They will be familiar to scholars of Kant, who famously argued that

> Man cannot dispose over himself because he is not a thing; he is not his own property; to say that he is would be self-contradictory; for in so far as he is a person he is a Subject in whom the ownership of things can be vested, and if he were his own property, he would be a thing over which he could have ownership. But a person cannot be a property and so cannot be a thing which can be owned, for it is impossible to be a person and a thing, the proprietor and the property. (1963, p. 165)

Why, exactly, did Kant hold that we cannot own ourselves? The passage quoted above suggests that he reasoned as follows: our self is a person, persons cannot be 'things', and since only things can be owned, persons therefore cannot be owned. But why not allow that we can be both persons and things, and therefore ownable? Kant owes us an argument for the view that personhood is incompatible with being an ownable thing.

One argument that he provided invoked the apparent obligation not to destroy oneself:

> An external object which in terms of its substance belongs to someone is his *property* (*dominium*), in which all rights in this thing inhere (as accidents of a substance) and which the owner (*dominus*) can, accordingly, dispose of as he pleases (*ius disponendi de re sua*). But from this it follows that an object of this sort can be only a corporeal thing (to which one has no obligation). So someone can be his own master (*sui iuris*) but cannot be the owner *of himself* (*sui dominus*) (cannot dispose of himself as he pleases)—still less can he dispose of others as he pleases—since he is accountable to the humanity in his own person. (1996, p. 6, 270, emphasis in original)

Kant's argument here is founded on two premises: (i) that one is obliged not to destroy a person, and (ii) that one is *not* obliged not to destroy one's

property. But both of these premises are open to question. In what follows, I focus on (ii).[14]

Note first that (ii) is clearly implausible if the obligation in question is understood broadly such that the claim is that there is no moral obligation *of any kind* not to destroy one's property. It seems clear that there can be an obligation not to destroy something one owns. Suppose that you are impoverished. Jobs are scarce, and the low-paying job that you have is the only thing preventing you and your children from starving. Your trusty bicycle, on which you commute to said job, is essential to your keeping this job. In this case, it seems that you have an obligation not to destroy your bicycle, though it surely remains your property. This suggests that having a property right over something is consistent with having an obligation not to destroy it.

Kant presumably foresaw problems like this. He could avoid them by understanding the obligation in question more narrowly. Perhaps he meant only to claim that there is no obligation *owed to one's property itself* not to destroy it, whereas there is an obligation *owed to a person herself* not to destroy her.[15] Note, however, that even on this interpretation, the claim about property is questionable. Many people think that sentient animals can be owned, yet there is, very plausibly, an obligation, owed to such an animal, not to destroy it. Many people also think that one can own a family heirloom or a historically significant artwork or a culturally significant tract of land, yet there is very plausibly an obligation not to destroy such a thing, and it is at least arguable that this obligation is owed to the thing itself.

More importantly, even if, for all of the clear examples of property, there is no obligation, owed to the thing itself, not to destroy it, it is not clear why we should think that this is a necessary feature of property. Indeed, there is, I think, some evidence that it is in fact *not* a necessary feature: we can clear-headedly imagine that something (a relatively sophisticated sentient animal, say) might be both the object of an obligation, owed to that thing, not to destroy it, and also someone's property. It may be debatable whether any actual entities fit this description, but thinking that there are such beings does not, or at least does not clearly, imply a conceptual confusion.

Are there other grounds on which one might reject the self-ownership thesis? There are at least three.

First, one might advert to the *non-transferability* of one's control over oneself. Paradigmatic objects of ownership (such as land and physical objects

[14] For an argument against (ii) that is similar to mine, see Taylor (2004, p. 67).
[15] I thank Jonas Hertel for pressing me to consider this formulation.

external to the body) are such that it is physically possible for the property (that which is owned) to be fully transferred into the control of others. One person can take control of another person's land, or of objects like buildings, bicycles, and bricks. It might be suggested that this transferability is in fact a necessary feature of property,[16] and that the self, not being transferable in this way, is therefore ineligible to be property.

Second, one might advert to the non-transferability not of the self, but of our *rights over* the self. Paradigmatic property rights can be fully transferred to others. We can fully transfer our property rights over a building, a bicycle, or a brick by giving or selling that item of property to another person. It might be thought that this transferability of property rights is a necessary feature of them.[17] And it might be thought that this precludes the self from qualifying as property, since our rights over ourselves cannot be fully transferred to others. (This is, I think, part of the thought behind the view, which most of us find compelling, that one cannot sell oneself into slavery.)

Third, one might advert to the fact that, where a person does own something, the self is—or partly comprises, or is partly comprised by—that which does the owning.[18] Thus, to accept self-ownership is to accept that an owner can own (part of) itself. This, it might be thought, is implausible or perhaps even conceptually confused: perhaps the concept of ownership presupposes that the owner and her property are non-overlapping.

I believe that none of these three appeals is compelling. Though I concede that both paradigmatic forms of property, and our property rights over it, are transferable, it seems to me doubtful that either type of transferability is a necessary feature of property. (I should emphasise here that I am interested in moral, not legal property; the transferability of *legal* property and *legal* property rights is not relevant to my purposes.) Similarly, though I concede that paradigmatic forms of property are external to the owner, I see no reason to suppose that it is a necessary feature of property that it be such.

[16] James Penner (2000, p. 78) comes close to making this suggestion in relation to legal property: 'If a person is unable to free himself of a thing, then to that extent he is controlled by his relationship to the thing, if not by the thing itself. Such control is antithetical to the idea of property.'

[17] Christopher Essert (2016, p. 267) takes it to be an invariable feature of property rights that they are 'at least in principle, alienable'. Meanwhile, Penner (2000, p. 80, see also pp. 81–87) argues that 'the right to transfer property is an inherent feature of property rights'.

[18] Similar objections to have been raised against the claim that we enjoy legal property rights over our bodies. See, for example, Penner (2000, Ch. 5) and S. Douglas (2011, p. 15). In *R. v Bentham* (2005 UKHL 18), Lord Bingham held that '[o]ne cannot possess something which is not separate and distinct from oneself. An unsevered hand or finger is part of oneself. Therefore, one cannot possess it . . .'.

Consider first the transferability of property itself. It is possible to concoct cases in which something is or becomes non-transferable, yet it seems to remain property. Suppose, for instance, that your hat becomes attached to your head in such a way that it can no longer be removed without it disintegrating into dust. It seems to me clear that this hat remains your property though it is no longer possible for you to transfer it into the control of another. It might be objected that it is still *logically* possible for you to transfer your hat to the control of another. Perhaps that is all that is required of property. However, it is, I think, also logically possible to transfer oneself into the control of another: indeed, we can imagine sci-fi scenarios in which a sophisticated scientist can control everything going on in one's body and mind by means of electrical currents and magnetic fields. So if logical transferability is all that is required, the requirement will give us no basis for denying self-ownership.

We can respond to the appeal to the transferability of property *rights* in a similar way. It is not at all clear that such transferability is a necessary feature of property. Indeed, there are, plausibly, instances of property rights that cannot be transferred to others. Suppose that I own some object—a vintage pocket watch, say. I then offer to transfer all of my rights over this watch to you, with the exception of the right to further transfer these rights; *this* right, I specify, will, if you accept my offer, henceforth be possessed by no-one. Suppose you accept my offer and thus end up with all but one of the 'subrights' that are characteristic of property rights. You possess the right to control the watch, to use it, against interference with it, and so on. But suppose that, as per our agreement, you lack the right to transfer these rights to others. No-one now possesses that right. Thus, the other rights have become non-transferable. It seems to me that in this case it would nevertheless be apt, or at least not obviously inapt, to describe you as owning the pocket watch. This casts doubt on the thought that property rights must be transferable.

What of the third appeal—the appeal to the claim that owners cannot own (parts of) themselves? I confess that, once we allow for property rights that cannot be transferred in full to others, I see little to be said in favour of this claim. Though ownership is often, and perhaps paradigmatically, a relationship that obtains between a person and objects that are entirely external to the person, I see nothing in the concept to rule out ownership of parts of one's person.[19]

[19] Here, I follow G. A. Cohen (1995, pp. 210–13).

I suspect the thought that owners cannot own (parts of) themselves gleans its initial air of plausibility from the thought that a person cannot sell herself into slavery. One possible explanation for this is that (i) even an enslaved person could plausibly own things, such as external possessions, so enslaved persons would still count as 'owners', but (ii) owners cannot be owned. However, there is another explanation available: we already own ourselves, and our property rights over ourselves cannot be transferred in full to others. Thus, we cannot make ourselves slaves. Given the availability of this alternative explanation, there is no need to suppose that owners cannot be owned; we can already rule out the possibility of selling ourselves into slavery.

Each of the arguments against the self-ownership thesis that I have considered in this section has had the same basic form. Each points to some feature of paradigmatic property rights, holds this feature to be a *necessary* feature of property rights, and maintains that our rights over our bodies lack that feature. In each case, I have responded by arguing that the putative necessary feature of property rights is not in fact a necessary feature, so does not prevent bodily rights from qualifying as property rights. Note, however, that even if some of the features that we have considered *are* in fact necessary features of property, and that our rights over our bodies are therefore not property rights, it remains possible—and I think plausible—that our rights over our bodies are in many important respects like property rights. Perhaps, indeed, it will still be plausible to maintain that our rights over our bodies are 'property*' rights, where property* rights are in all important respects like property rights, except in the respects discussed in this section. And this, I think, would be sufficient to sustain a variant of the argument from self-ownership—a variant that is formulated in terms of property* rights rather than property rights. Nevertheless, in what follows, I will continue to formulate the argument from self-ownership in terms of ordinary property rights.

III. Property Rights Protect Only Against Physical Interferences

Recall that the argument from self-ownership runs as follows:

[1] We own ourselves.
[2] If we own ourselves, then we own our minds.
[3] We own our minds (from [1] and [2]).

[4] If we own our minds, then we possess a right against interference with our minds.

[5] We possess a right against interference with our minds (from [3] and [4]).

Note that, while this is an argument for *a* right against mental interference, it is not clearly an argument for *the* right against mental interference that I have been characterising in earlier chapters. It could be that the right asserted in [5]—and supported by [1]–[4]—is an altogether different sort of right than *the* right against mental interference that I have been characterising. (I henceforth use the abbreviation 'RAMI' to refer only to the right characterised in earlier chapters.)

In this section, I consider an objection that capitalises on this possibility. It maintains that the right established by the argument from self-ownership is more limited than the RAMI. More specifically, it maintains that the argument from self-ownership can establish only a right against *purely physical* interferences with the mind, whereas the RAMI provides broader protection.

As I characterised the RAMI in earlier chapters, it protects not only against mental alterations produced through a purely physical (i.e. not-at-all mental) process in the influencee—as when the alteration is produced via the administration of drugs, electric currents, or magnetic fields—but also against alterations produced via a (partly) mental process in the influencee, where that process is insensitive to the reasons that bear on the alteration. Examples of the latter would plausibly include causing feelings of disgust through the use of subliminal imagery, inducing seizures through flashing lights,[20] and attenuating a person's gloominess via hypnosis.[21]

It might be thought, however, that the rights against interference entailed by ownership are more modest—they are rights against some more narrowly specified form of interference. Some scholars have argued, for example, that the interference prohibited by property rights must be *physical* interference. For instance, Robert S. Taylor holds that 'the incidents (claim-rights) that constitute self-ownership impose perfect duties of *physical* noninterference upon others—that is all' (2004, p. 67 (my emphasis)).[22] It is not clear exactly

[20] I assume here that perception is an at least partially mental process.

[21] In Chapters 6 and 7, I will consider, in much more detail, which mental influences infringe the RAMI.

[22] Taylor is here referring to *moral* property rights over the *self*, but similar claims are sometimes made in relation to legal property rights over external objects. See, for example, S. Douglas (2011, pp. 17–18).

what these scholars think would count as *non-physical* interference, but a plausible thought is that alterations that are produced via partly mental processes within the entity being interfered with would qualify as non-physical.

Thus, whereas on my view, intentionally diminishing someone's gloominess through a nonconsensual hypnotic procedure interferes with her mind, on the understanding of 'interference' relevant to property rights, perhaps it does not.

Arguably, then, we can establish only part of the RAMI by appealing to a property right over the mind; we can establish only the part of the RAMI that protects against mental alterations produced via purely physical, that is, not (partly) mental processes.

We could evade this objection if we could reject the view that the interference prohibited by property rights (henceforth, property-relevant interference) must be purely physical—if, more precisely, we could reject what I will call 'the physicality requirement':

> Intentionally altering the state of a person's property constitutes property-relevant interference only if the influencer produces the alteration via a purely physical—that is, not (partly) mental—process within the owned entity.

I cannot offer a knock-down argument against the physicality requirement, but I think there are two considerations that tell against it.

First, it is, I think, possible to identify cases in which mentally mediated alterations with an item of property plausibly qualify as property-relevant interferences. Suppose Farmer Brown owns a cow, Daisy, which one of her neighbours, Farmer Green, wants to kill. One night, Farmer Green hypnotises Daisy, and induces her to walk to her death into a deep river. At no point in this process does Farmer Green touch Daisy.[23]

It seems to me that, in addition to wronging Daisy, Farmer Green wrongs Farmer Brown. And it seems to me that one of the wrongs that Farmer Green commits against Farmer Brown is the wrong of interfering with her property—namely, Daisy. Yet the alterations that Farmer Green causes to Daisy in this case are caused by a partly mental process within Daisy.

[23] This case is inspired by two seventeenth-century English legal cases, *Farmer v Hunt* and *Durant v Childe*, in which the defendants chased the claimant's cattle, causing them to die, but without touching them, and were found liable for trespass against property. See, for discussion, S. Douglas (2011, p. 111).

Second, the physicality requirement seems morally arbitrary. It is simply not clear *why* the mere involvement of mental processes should render an influence on a person's property less problematic, morally speaking. The involvement of mental processes does not, in itself, seem morally significant.

Of course, it is one thing to say that mentally mediated influences on property *sometimes* infringe the property right; it is another to say that they do so whenever they meet the precise conditions that I outlined when characterising the RAMI in Chapter 2. For all I have said in this section, it could be that the conditions for property-relevant interference remain tighter than the conditions for RAMI-infringing interference. The argument from self-ownership thus falls short of establishing a right against mental interference *precisely as I have specified that right*. However, in rejecting the physicality requirement, we have, I think, rejected the most obvious basis for claiming that the conditions for property-relevant interference are tighter than the conditions for RAMI-infringing interference.

IV. Superficiality

A third objection to the argument from self-ownership holds that it is problematically superficial. The argument seeks to derive the RAMI from a property right in the self that is arguably no more fundamental, ontologically, than the RAMI. Indeed, since the RAMI is plausibly a *constituent* of self-ownership—one of the bundle of rights that comprise it—one might think that the property right is the *less* fundamental right; the RAMI partially grounds self-ownership, not the reverse. Thus, it might be thought that the argument gets the order of justification wrong. Rather than defending the RAMI by appealing to self-ownership, we should defend self-ownership by, *inter alia*, defending the RAMI that partly constitutes it.

I resist the idea that justifications must run from more to less ontologically fundamental. One can sometimes legitimately defend claims about something more fundamental by appealing to claims about something less fundamental. For example, one can support claims about neural activity ('she must have injured part of her brain') by reference to claims about mental activity ('she was unconscious after the fall') and one can support claims about smaller physical entities ('there are buildings on the banks of the Tyne') by reference to larger entities that they partly comprise ('there's a city on the banks of the Tyne').

Problems arise, however, when the less fundamental claim is also epistemically redundant in the sense that it provides no evidence for the more fundamental claim beyond that which is already provided by our evidence for the less fundamental claim. In such cases, we can safely eliminate the less fundamental claim from our justification, moving straight from our evidence for it to the more fundamental claim. Consider a case in which we seek to support the claim that April must have injured part of her brain by claiming that she was unconscious after the fall. But suppose that the only evidence we have for the view that April was unconscious is that we have seen a scan of her brain showing damage of a sort that typically causes unconsciousness. In this case, the claim about April's unconsciousness is epistemically redundant. We could reason directly from the brain scan result to the claim that April has injured part of her brain. Her unconsciousness can be eliminated from the justification of that claim.

In the case of interest to us here, the problem would arise if coming to know that we own ourselves would provide no additional evidence for the claim that we possess the RAMI beyond that which is already provided by our *evidence for* the self-ownership thesis. We could then reason directly from the evidence for self-ownership to the RAMI without having to reason via self-ownership.

Is the self-ownership thesis epistemically redundant in this way? It could be. Suppose our property right over ourselves is just a bundle of fully independent rights—by which I mean rights with different grounds and different evidential support. And suppose the RAMI is one of the members of this bundle. If so, then the self-ownership thesis would be downstream, in evidential terms, from the RAMI. All of our evidence for self-ownership would either support self-ownership *by* supporting (perhaps among other elements in the rights-bundle) the RAMI, or would not at all support the RAMI, since it is only evidence for other, independent elements in the rights-bundle.

So I concede that, in some evidential contexts, the argument from self-ownership will be epistemically redundant. However, I see no reason to suppose that this will *always* be so. Suppose, for example, that our evidence for the self-ownership thesis is that it best explains how we came to acquire external property. Or suppose we simply think that the self-ownership thesis is self-evident. Or suppose we accept the self-ownership thesis on the basis that, taken together, the bundle of rights which comprise self-ownership help to protect the owner's well-being or autonomy. In these circumstances, it is not clear that our evidence for the self-ownership thesis would directly (i.e. independently of supporting self-ownership) support the RAMI.

So I deny that the argument from self-ownership will always be epistemically redundant. Moreover, even in contexts where the argument *is* epistemically redundant, it may still serve an important *dialectic* function. It may be that the self-ownership thesis provides no evidence for the existence of the RAMI beyond that already provided by our evidence for the self-ownership thesis. Thus, it may be that, in offering an argument from self-ownership to the RAMI, we are providing no new justificatory support for the RAMI. Still, this argument may help to convince people to accept that we possess the RAMI. This is especially likely if people tend to find the self-ownership thesis more plausible or more salient than the evidence for it.

V. Conclusion

In this chapter, I have offered a theoretical defence of the RAMI: the argument from self-ownership. I first laid out the argument from self-ownership and motivated its three premises. I then considered three objections to the argument. These appealed respectively to the claims that (i) we do not own ourselves, (ii) a right against mental interference implied by self-ownership would be narrower in scope than the RAMI, as I have characterised it, and (iii) the argument is problematically superficial. I sought to show that none of these objections undermines the argument. At least in certain evidential contexts, the argument does provide some additional justificatory support for the RAMI, and even in other contexts, it may have some dialectic force.

6

The Scope of the Right Against Mental Interference*

In Chapter 3, I gave a rough, preliminary account of the scope of the right against mental interference (RAMI)—of which actions infringe it. In this chapter, I further develop that account.

Recall that, on my account, *A* interferes with *B*'s mind if

> *A*'s action *X* alters a mental state of *B*'s (the *agential alteration condition*)

and

> the mediating process—the process within *B* via which *X* produces the alteration—is insensitive to the reasons that bear on the alteration (the *insensitive process condition*)

where the 'reasons that bear on the alteration' refers to the reasons for or against *B*'s bringing about or undergoing the alteration.

The agential alteration condition and insensitive process condition are, I claim, jointly sufficient for *A* to mentally interfere with *B*. Three further conditions are jointly sufficient for this mental interference to infringe *B*'s RAMI. *A*'s mental interference with *B* infringes *B*'s RAMI if

> *A* is a well-situated moral agent (the *duty-bearer condition*)[1]

* I thank Alex Kaiserman and Gabriel De Marco for very helpful discussions regarding topics addressed in this chapter.

[1] Recall that I am taking the class 'moral agents' to include at least all adult human beings who are not severely cognitively or morally impaired, and am assuming that these beings are also 'well-situated' if (i) they could reasonably be expected to foresee the consequences or other features of their conduct by virtue of which they interfere with the putative right-holder's mind in a particular case, and (ii) there is nothing severely impeding their exercise of their cognitive and moral capacities.

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0006

and

> *A*'s interference with *B*'s mind is significant (the *significance condition*)

and

> *B* possesses the RAMI and has not waived or forfeited it in relation to the interference in question (the *nonliability condition*).

In this chapter, I further spell out the scope of the RAMI by developing what I take to be the two most contentious and unclear of the five abovementioned conditions: the significance condition and the insensitive process condition. I consider the significance condition in Section I, and the insensitive process condition in Section II. In the next chapter, I will make the account of my scope of the RAMI more concrete by assessing its implications for some specific types of influence, such as neurointerventions, nudging, aversion therapy, and deception.

I. The Significance Condition

According to the significance condition, interference by a well-situated moral agent with the mind of a nonliable individual infringes the RAMI *if* the interference is significant. I will take the significance of a mental interference to be a function of the magnitude of the interference; a mental interference is significant if and only if its magnitude lies above some threshold. I cannot here precisely specify the location of this threshold, and I also cannot offer a full account of what determines the magnitude of a mental interference, however, in what follows, I will suggest a number of factors that I think plausibly play a role in determining this magnitude. The relevance of each of these factors is, I claim, supported by common-sense thinking about mental and/or bodily interference. However, I do not claim that these factors are the *only* determinants of the magnitude of an interference.

A. Importance of Implicated Mental States

The first factor that, I suggest, contributes to the magnitude of a mental interference is the importance of the mental states implicated in the interference;

other things being equal, interfering with more important mental states will constitute a greater interference than interfering with less important mental states.

The relevance of this factor is supported by common-sense judgments about bodily interference. It is, for example, somewhat plausible to think that excising a cubic centimetre of brain or heart tissue constitutes a greater bodily interference than excising a cubic centimetre of calf muscle.[2] Similarly, it is plausible to think that the magnitude of a mental interference will depend on the importance of the mental states interfered with such that, for example, altering a person's religious commitments or intimate attachments constitutes a greater interference than altering her ranking of ice-cream flavours.

How are we to determine the importance of a mental state? A significant factor here may be the degree to which the mental state in question contributes to sustaining our other mental states or, to approach things from the opposite direction, the degree to which other mental states depend on the mental state in question. ('Contributes' and 'depend' could be understood as referring to a causal relationship, a constitutive relationship, a normative relationship, or some combination of these.) Thus, for example, consider a person's implicit acceptance of basic logical principles, a commitment to the scientific method as a reliable guide to the truth, and the belief that everything happens because God wills that it is so. These would likely count as important mental states, and they may do so at least in part because they play a role in sustaining a great many of one's other mental states. By contrast, consider my intention to keep writing for another five minutes, my fleeting feeling of annoyance at the noise from a passing truck, or the brief awareness of a bird pecking at the ground in my peripheral vision. These mental states likely do not count as highly important, and one reason for this may be that few other mental states depend on them.

However, other factors may also be relevant to importance. For example, the importance that we *ascribe* to a mental state may be relevant to its actual importance. Perhaps the greatest mental interferences that we can imagine would involve altering our religious, philosophical, and political commitments as well as our attachments to significant others. It is plausible that altering such commitments and attachments can constitute a very

[2] This and all other similar claims in this section are intended to be read as containing an implicit *ceteris paribus* clause; *other things being equal*, excising a cubic centimetre of brain or heart tissue interferes with a person's body to a greater degree than does excising a cubic centimetre of calf muscle.

great mental interference in part because these are very important mental states. Why are they important? Perhaps because many other mental states depend on them. But suppose that it turns out that this is not so. Suppose that my philosophical commitments and my attachment to, say, some long deceased friend, actually play very little role in the rest of my mental life. Nevertheless, it seems that these could be important mental states. I propose that they could be important simply because I, in some way or other, ascribe them great importance. Perhaps I strongly identify with them, or regard them as central to my life narrative. Or perhaps I simply strongly value them.

There is a question about whether, in addition to the importance of the mental states implicated in a mental interference—that is, the importance of the mental states that are interfered with—the magnitude of a mental interference might also be affected by the importance of any mental states indirectly affected by the interference. Should we take into account mental alterations to which the interfering act contributes, but which do not count as part of the interference, for example, because the influencing act makes only a minor or distant causal contribution to the alteration, or the alteration is brought about via a reasons-sensitive process? Recall *Brain Drug* and its variants, introduced in Chapter 4. In these cases, you make a major and proximate contribution to an alteration in Lucien's level of gloom, and you do so via a reasons-insensitive process in Lucien. Thus, the alteration to Lucien's level of gloom is partly constitutive of the mental interference. But suppose the reduction in Lucien's level of gloom also causes further mental effects. Suppose it causes him to develop a new passion for model trains. *This* mental alteration may be produced via a reasons-sensitive process. Or it may be one to which your action makes only a minor or distant causal contribution. Either way, it need not be part of the mental interference that you perpetrate, and there is a question about whether it should be taken into account in determining the magnitude of your interference.

In the case of bodily interference, we would not generally consider the magnitude of an interference to be affected by bodily alterations to which the original interfering act made only a minor or distant causal contribution. Suppose two individuals undergo a very similar operation—an appendicectomy, say. In one of these individuals, the operation contributes, alongside various other factors such as the patient's genetic susceptibility to infections and a subsequent injury to the surgical scar, to an infection in the site of the scar several years later. In the other individual, no such complication occurs.

I doubt we would want to say that this difference in long-term outcomes makes any difference to the magnitude of the bodily interference involved in these two operations. Similarly, I doubt we would think it relevant to the magnitude of interference if, in one case, the operation leads, via a reasons-sensitive process, to further bodily alterations. Suppose that one of the patients, but not the other, takes a strong painkiller as a result of deliberating about how best to control her post-operative pain. The bodily alterations produced by this painkiller would not clearly affect the degree of bodily interference involved in the operation.

So I suspect that, while mental states that are partly constitutive of a mental interference contribute to the magnitude of the interference, mental states affected by knock-on alterations do not. Still, I do not exclude the possibility that the latter mental states—or some subset thereof—might make some contribution.

B. The Number of Mental States Implicated

A second factor that plausibly contributes to the magnitude of a mental interference is the number of mental states that are implicated in the interference.[3] This factor again has a bodily analogue. Amputating a person's foot and leg interferes with a person's body to a greater degree than does amputating her foot alone. Breaking all of a person's limbs interferes with his body to a greater degree than does breaking only one of his limbs. Similarly, it is plausible that the magnitude of mental interference depends on the number of mental states implicated in the interference. Deleting all of a person's memories interferes with her mind to a greater degree than does deleting only one memory. Subliminally inducing a hatred for all shades of pink interferes with a person's mind to a greater degree than subliminally inducing a hatred for one very specific shade of pink.

There will of course be very difficult questions regarding how we should individuate mental states for the purposes of counting them. But in at least some cases—like those just mentioned—it seems possible to compare two mental interferences and say that one interferes with more mental states than the other. Moreover, differences in how we individuate mental states may not

[3] Again, I remain open to the possibility that mental states that are altered through knock-on effects of the interference may also be relevant here, but I focus on mental states that are implicated in (i.e. partly constitutive of) the mental interference.

make much difference to overall assessments of the magnitude of an interference. The more finely we individuate mental states, the greater the number of mental states that will be interfered with in a particular case, but also the less important those mental states will be. The more coarsely we individuate mental states, the smaller the number of mental states that will be interfered with in a particular case, but also the more important those mental states will be. Depending on how we balance number and importance against one another, these effects may offset one another, leaving the overall magnitude of the interference unaffected.

C. Degree of Alteration

Another factor that plausibly contributes to the magnitude of a mental interference is the degree of change to the person's altered mental states.[4]

Again, the relevance of this factor is suggested by common-sense thinking about bodily interference. Where a bodily interference consists in altering a person's bodily states, the magnitude of the bodily interference is affected by the degree of change to the person's bodily states. Thus, cooling a person's feet down by 20 degrees constitutes a greater bodily interference than cooling a person's feet down by one degree. Similarly, destroying a person's whole liver constitutes a greater bodily interference than mildly damaging a person's whole liver. It is, I think, similarly plausible that the degree of change to a person's mental states will affect the magnitude of a mental interference. Thus, for example, causing a person who previously liked Vegemite to hate it constitutes a greater mental interference than causing her to like it slightly less. Similarly, causing someone who (correctly) believes that 'Scafell Pike' refers to a mountain in England to believe that it refers to a children's book character constitutes a greater mental interference than causing her to believe that it refers to a mountain in Wales.

What determines the *degree* of alteration to a mental state? I believe that—at least for many types of mental state—two factors will be relevant: how much the *strength* of the mental state is altered, and how much its *content* is altered.

[4] Again, I remain open to the possibility that the range of mental alterations relevant here may be wider than the range that are partly constitutive of the mental interference.

What determines the strength of a mental state will depend on the type of mental state in question. Some mental states, such as likes and wants, intrinsically admit of degrees; for any like or want, it makes sense to ask 'how much do you like *x*?' or 'how much do you want *y*?' Other kinds of mental states, such as beliefs, are intrinsically binary; you either believe that the world is flat or you do not. However, even for mental states of this kind, we can often identify a plausible metric of strength. For example, we could think of the credence that you attach to the world's not being flat as a metric of the strength of your belief that the world is not flat.

What determines how much the *content* of a mental state has been changed will also depend on the nature of the mental state. Consider two claims: the experience of redness is more similar, in its content, to the experience of orangeness than it is to the experience of blueness; the belief that 'Scafell Pike' refers to a mountain in England is more similar to the belief that it refers to a mountain in Wales than it is to the belief that it refers to a children's book character. Each of these claims is plausibly correct; however, the metric of similarity that explains their correctness is not the same. The first claim is correct in virtue of the phenomenology of redness, orangeness, and blueness. The second claim is correct in virtue of some facts about geography and fiction.

I am sorry to say that I cannot develop an account of similarity of content or strength for even one type of mental state—let alone for all types—so I am simply going to leave it as an intuitive matter. Fortunately, we do, I think, generally have a good intuitive grasp of both similarity in strength and similarity in content with respect to a wide range of mental states.

D. Concluding Comments on Magnitude

The thoughts in this section have been sketchy. But I hope they give an idea of the kinds of considerations that may affect the magnitude of a mental interference.

There is, of course, also the further question of how great a magnitude a mental interference must have in order to count as a *significant* interference, and thus one that might meet my jointly sufficient conditions for infringement of the RAMI. I am going to leave this question open. However, in the next chapter, I will offer some comments on how the location of this threshold could affect the protection provided by the RAMI against various prevailing forms of mental influence.

II. The Insensitive Process Condition

The significance condition—discussed in the previous section—plays a role in determining whether a mental interference infringes the RAMI. On my view, a mental interference infringes the RAMI whenever it meets the significance condition and two other conditions: the duty-bearer condition and the nonliability condition.

In the remainder of the chapter, I will turn my attention to the conditions for an action to count as a mental interference. Recall that these are the agential alteration condition and the insensitive process condition. My focus will be on the latter condition, which I take to be the less straightforward of the two. This condition picks out cases in which the mediating process—the process within the influencee via which the alteration is produced—is insensitive to the reasons that bear on the alteration.

As I have presented my account of mental interference thus far, the agential alteration condition and insensitive process condition are jointly sufficient for mental interference (I do not claim that either is necessary for it). However, in what follows, I will often present the insensitive process condition as alone sufficient. This is because I will be assuming that the agential alteration condition is satisfied—that we are considering actions that alter another's mental states. I will continue to use the term 'mental influence(s)' to refer to such actions.

My question in this section will be, 'what does it take for a mediating process to be insensitive to the reasons that bear on the alteration?' or, as I will henceforth sometimes abbreviate this, 'what does it take for a mediating process to be reasons-insensitive?'.

In Subsection II.A, I introduce some cases that will serve as a starting point for the subsequent discussion. In Subsection II.B, I outline my approach to developing a sufficient condition for reasons-sensitivity. In Subsection II.C, I state my sufficient condition in preliminary form. In Subsection II.D, I refine this condition so as to capture a broader range of influences. And in Subsection II.E, I consider and reject a proposal for broadening it further.

Before embarking on these tasks, however, I need to introduce an important clarification. To this point, I have been writing as though reasons-sensitivity is an all or nothing matter—as though a mediating process is either reasons-sensitive or not. But in fact, I think reasons-sensitivity comes in degrees. The process mediating a mental influence can be highly reasons-sensitive, somewhat reasons-sensitive, or only minimally reasons-sensitive, for example. Or it can be entirely insensitive to reasons. I do not really want

my insensitive process condition to pick out only mediating processes that are *entirely* insensitive to reasons. Rather I want it to pick out those mental influences in which the reason-sensitivity of the mediating process *is not to a significant degree* reasons-sensitive, where that implies that the process' level of reasons-sensitivity lies below some rather low threshold. Thus, I will henceforth understand the insensitive process condition as holding that a mental influence constitutes a mental interference if the mediating process is *not significantly* sensitive to the reasons that bear on the alteration, and I will use the unqualified term 'reasons-insensitive' to mean 'not significantly reasons-sensitive'. When I wish to refer to a process that is *entirely* reasons-insensitive, I will make that explicit.

A fully satisfying elucidation of the insensitive process condition would answer two fundamental questions: What determines the reasons-sensitivity of the mediating process? And where does the threshold for (in)significant reasons-sensitivity lie? But both of these questions are difficult and any answer that I could plausibly offer to them here would, I think, be highly contestable. Thus, in line with the approach taken elsewhere in this book, I will remain noncommittal on both questions (though I will, later in this section, schematically trace out some approaches that might be taken to answering the second). Instead, I will offer what I hope will be a relatively uncontroversial sufficient condition for a mediating process to count as reasons-insensitive.

In working my way towards this condition, I will continue to make the same assumptions about reasons that I made in Chapters 2 and 3. I will understand reasons to be considerations (and more specifically, facts) that count for or against something, either absolutely or relative to the situation in which the fact does not obtain. I take 'the reasons that bear on an alteration' to be the reasons for or against the influencee undergoing or bringing about the alteration, and I allow that these reasons may be epistemically accessible or inaccessible for the influencee. I also allow that a person may have both preference-based reasons (fundamental reasons that are facts about the person's preferences, and derivative reasons that derive from such fundamental reasons) and non-preference-based reasons (fundamental reasons that are not facts about the person's preferences, and derivative reasons that derive from such fundamental reasons). Facts that could plausibly serve as fundamental non-preference-based reasons bearing on a mental alteration would plausibly include (i) facts about the 'fit' between the implicated (pre- and post-alteration) mental states and the external world, (ii) facts about the intrinsic value of the implicated mental states (perhaps including facts about

their pleasurability or painfulness), and (iii) facts about the instrumental value of the implicated mental states.

A. Paradigmatic Cases

Before proceeding to offer my sufficient condition for reasons-insensitivity, it will be helpful to introduce some paradigmatic cases. Consider first:

> *Rain Claim*. Ada arrives late to a meeting in a windowless room and says to all the attendees, as she enters the room, 'Gosh, it's raining cats and dogs out there'. She does this with the intention of communicating to those in the meeting that it is raining, as indeed it is. Her utterance induces Bert, one of the attendees, to think about whether Ada's claim is likely to be true. For example, Bert reflects on Ada's character and comes to the (let us suppose correct) conclusion that Ada is a trustworthy individual. Bert also recognises that there is some other evidence that supports Ada's claim. For example, he notes that Ada's clothes appear to be soaked, recalls that the generally reliable weather app on his phone reported a 93% chance of rain today, and reflects on the fact that he had seen dark clouds gathering when he arrived at the meeting himself, fifteen minutes previously. All in all, Bert concludes, the evidence overwhelmingly supports Ada's claim that it is raining. Thus, whereas he had, before Ada had entered the room, not believed that it was raining, now he forms the belief that it is.

Ada exerts a range of mental influences on Bert in this case—her actions produce various different alterations in Bert's mental states. But let us focus on one particular mental influence that she exerts: the influence on his beliefs about whether it is raining. This seems to be a clear case of a mental influence that *does not* satisfy the insensitive process condition; the process within Bert via which Ada's action alters Bert's beliefs appears to be significantly sensitive to at least some reasons that bear on the alteration, such as the fact that a somewhat reliable source has reported that it is raining, the fact that a reliable weather app reported a high chance of rain, and the fact that dark clouds were gathering a short time previously. Indeed, the process via which Bert arrives at his belief-change is a paradigm example of a highly reasons-sensitive process.

In *Rain Claim*, Bert comes to believe that it is raining through consciously reflecting on the reasons that bear on this alteration. But a mediating process

can be significantly sensitive to reasons without involving anything like conscious reflection on reasons. Consider:

> *Rain Gesture*. Ada arrives late to a meeting in a windowless room, drenched from the rain outside. As she enters the room, she gestures towards her soaked clothing with the intention of communicating to the attendees that it is raining outside. Bert, whose reliable weather app had reported a 93% chance of rain today, who had seen dark clouds gathering when he entered the building a short time earlier, and who knows that Ada is a trustworthy individual, sees Ada's gesture. Without any mediating conscious thought, he comes to believe that it is raining outside, whereas he did not previously believe this.[5]

The process within Bert via which Ada's gesture produces the belief-alteration in this case is a rapid, subconscious one. Nevertheless, it is plausible to suppose that it is significantly reasons-sensitive. One reason that Bert has to believe that it is raining in this case is that a trustworthy source has expressed—albeit implicitly—that it is raining. And Bert is plausibly somewhat sensitive to this reason. We can see this by reflecting on what might have happened had the reason been absent or weaker. Had Bert been aware that Ada was in fact known for weather-related trickery, he would plausibly have been at least somewhat less disposed to change his belief, and this, I think, suggests that he was sensitive to the reason. Another reason that Bert has to change his belief is that his reliable weather app reported a high chance of rain. Again, Bert is plausibly somewhat sensitive to this reason. Suppose that Bert's weather app had reported only a 7% chance of rain. Again, he would then have been somewhat less inclined to form the belief that it is raining.

In both *Rain Claim* and *Rain Gesture*, then, the mediating process within Bert is plausibly significantly sensitive to the reasons that bear on the resulting alteration to his beliefs.

But consider now:

> *Neurostimulation*. While sitting in a boring meeting in a windowless room, the thought 'it's raining' pops into Bert's head, for no reason apparent to

[5] In offering a case in which both Ada's mode of communication and Bert's response to it are implicit, I am not meaning to suggest that these things always go together. We sometimes respond to explicit statements via implicit thought processes, and to implicit modes of communication via explicit thought processes.

> Bert, and this thought soon morphs into a full-blown belief, again, for no reason apparent to him; Bert simply finds himself with the—as it turns out, correct—belief that it is raining. In fact, unbeknownst to Bert, his hearing aid has been fitted with a sophisticated electrical brain stimulation device, and the belief was produced by a nefarious neuroscientist, Cindy, who is controlling this device, and who used it to modulate the patterns of neural activity on which Bert's belief that it's raining supervenes. Cindy regularly uses this technique to produce the belief that it's raining in people wearing hearing aids, and she does this regardless of whether it is in fact raining and regardless of what evidence exists on whether it is raining.

In this case, the 'mediating process' is the physico-chemical process within Bert via which Cindy's device alters Bert's beliefs. There is no reason to suppose that this process is at all sensitive to the reasons that bear on the alteration. Suppose that it had not been raining, and that there had been no evidence that it was raining. Presumably, Cindy would then have performed the same procedure, a similar process would have occurred within Bert, and this would have resulted in a similar change to his beliefs. This suggests that the process that produced his belief-change in the actual case was neither sensitive to the fact that it was raining, nor to the available evidence that it was raining. *Neurostimulation* is very plausibly a case in which the mediating process is not significantly sensitive to reasons. It is also a clear case of mental interference.

B. Counterfactual and Causal Approaches to Reasons-Sensitivity

Recall that we are interested in cases in which an influencer performs an action that alters the mental states of some influencee via a 'mediating process' within the influencee. Our question is: how are we to determine whether the mediating process is reasons-insensitive, that is, not significantly sensitive to the reasons that bear on this alteration? And our answer should conform to the intuitive verdicts that we made on the three cases discussed in the previous subsection: it should tell us that the mediating process is reasons-insensitive in *Neurostimulation*, but it should not tell us this about the mediating processes in *Rain Claim* and *Rain Gesture*.

In the philosophical literature, there are two main approaches to determining the degree to which something is sensitive to reasons. The first—the counterfactual approach—invokes the idea that we can determine whether

and to what degree something is sensitive to a reason or set of reasons by considering what happens in certain appropriately specified counterfactual scenarios in which the relevant reasons are otherwise.[6] Thus, for example, we might say that the performance of some action was sensitive to some reason that bears on that action to the extent that, across some specified set of counterfactual situations, whether the same kind of action occurs is correlated in the right way with the presence and strength of that reason.

The second approach—the causal approach—invokes the idea that for something to be sensitive to reasons is for it to occur *because of* reasons (Kaiserman 2021; Sartorio 2016, 2018).[7] For example, we might say that the performance of some action was sensitive to some reason to the extent that the reason causally contributed, in the right kind of way, to the action.[8]

In the previous section, I drew on counterfactual approaches. In discussing the three cases introduced there, I sometimes appealed to claims about what would have happened had the reasons present in the actual case been absent or weaker. For example, regarding *Neurostimulation*, I suggested that, had it not been raining, and had there been no evidence that it was raining, Bert would still have undergone the same change in his beliefs. And I took this to support the view that the mediating process that in fact altered his beliefs was reasons-insensitive. My underlying thought here was roughly as follows. If subtracting or weakening reasons that were in fact present would not have affected how likely the influencee was to undergo the alteration, then the process via which the alteration was produced was probably not significantly sensitive to reasons. On the other hand, if subtracting or weakening reasons that were in fact present *would* have affected how likely the influencee was to undergo the alteration, then the process via which the alteration was produced probably was significantly sensitive to reasons.

[6] For the classic statement of this kind of view, see Fischer and Ravizza (2000). Fischer and Ravizza's account does not allow for reasons-sensitivity (or, on their terminology, reasons-responsiveness) to come in degrees, but others have sought to develop their account in ways that would allow for degrees. See, for example, Coates and Swenson (2013) and Montminy and Tinney (2018).

[7] See also, for discussion of these approaches, Heering (2022).

[8] I will henceforth assume both that reasons are facts and that reasons can be causes. If you are sceptical that facts can be causes, please read my references to the causal influence of reasons as references to the causal influence of the events, states, or properties that are captured by the reason-fact. Consider a case in which the fact that the plane is experiencing turbulence is a reason for the pilot to lower the plane's altitude. If you are not convinced that this fact could cause the pilot to lower the plane's altitude, because you are not convinced that facts can be causes, please read the claim that this reason caused the pilot to lower the plane's altitude as the claim that the state of turbulence, turbulent movement of the plane, or the property of the air conditions of being turbulent caused the pilot to lower the plane's altitude.

I will continue to apply counterfactual tests of this kind as a rough heuristic—as an indicator of whether a mediating process is likely to be significantly sensitive to reasons. However, in seeking to develop a more precise test for reasons-insensitivity, I will take my inspiration not from counterfactual approaches, but from causal approaches. My reason for doing this has to do with my desire to be as ecumenical as possible. Counterfactual and causal approaches are both vulnerable to criticisms, but they are vulnerable to criticisms of different kinds. Counterfactual approaches are normally criticised for yielding implausible verdicts in certain cases—for example, cases in which reasons are, in the actual world, causally inert, though would have made a difference in certain counterfactual scenarios (see, e.g. Kaiserman 2021). By contrast, the main problem for causal approaches is not that they get the wrong results, but that they are insufficiently deep, since they employ concepts—like the concept of a cause—that themselves require further explication.

Given my goal of being ecumenical, it is more important that I offer a test that yields widely accepted verdicts than that I offer a test which provides great analytical depth, hence I offer a test inspired by causal approaches. I should be clear, however, that my test falls short of providing a full, causal account of reason-sensitivity. It is incomplete in two important ways. First, though it is informative on whether a mediating process is significantly sensitive to reasons or instead reasons-insensitive (i.e. not significantly sensitive to reasons), it is silent on differences in reasons-sensitivity above the threshold for significant reasons-sensitivity. And second, it offers only a sufficient condition for reasons-insensitivity, not a necessary one.

C. A Preliminary Sufficient Condition for Reasons-Insensitivity

In this subsection, I state my sufficient condition for reasons-insensitivity in preliminary form. I then, in the next subsection, revise it so that it can accommodate a wider range of cases. Here is my preliminary formulation. Suppose that an influencer performs an action that produces a mental alteration in an influencee via mediating process *P* within that influencee. My preliminary condition holds that *P* is reasons-insensitive if reasons bearing on the resulting alteration did not significantly causally contribute, via *P*, to the alteration's occurrence.

This preliminary condition requires some immediate glossing. I will assume that there are two ways in which reasons could causally contribute

to the occurrence of an alteration. They could positively causally contribute to it, for example, by promoting the alteration or making the alteration more likely to occur. Or they could negatively causally contribute to it, for example, by creating a barrier to the alteration or making it less likely to occur. In applying my preliminary sufficient condition, I will count reasons as making causal contributions to the occurrence of an alteration only when the reason contributed *in the right direction* to the occurrence of the alteration—reasons for the alteration positively contributed to the alteration, and reasons against the alteration negatively contributed to the alteration. Thus, the preliminary condition holds that mediating process *P* is reasons-insensitive if reasons bearing on the resulting alteration did not significantly causally contribute *in the right direction*, via *P*, to the alteration's occurrence. I will, however, henceforth generally omit reference to the direction of causal influences, taking it as read that causal contributions are in the right direction.

The preliminary condition, thus understood, gives plausible verdicts on the cases described above. The condition is not satisfied in *Rain Claim*, since Bert's formation of the belief that it is raining is caused, via his conscious reflection, in large part by reasons for that belief-change, such as the fact that a trustworthy source has reported that it is raining, the fact that the weather app reported a 93% chance of rain, and the fact that dark clouds were gathering a short time previously. Thus, my preliminary condition does not class Bert's process of conscious reflection as reasons-insensitive, and so does not imply that Ada interferes with Bert's mind in *Rain Claim.*

Similar thoughts apply to *Rain Gesture.* Here too, my preliminary condition is not satisfied. Reasons bearing on Bert's belief-change, such as the fact that a trustworthy source has implicitly expressed that it is raining, the fact that the weather app reported a 93% chance of rain, and the fact that dark clouds were gathering recently, have, plausibly, significantly causally contributed, via Ada's rapid, subconscious reaction to Ada's gesture, to Bert's change in belief.

By contrast, in *Neurostimulation*, my preliminary sufficient condition plausibly *is* satisfied, since there is no basis for supposing that reasons that bear on the alteration in this case are making any significant causal contribution to the alteration of Bert's beliefs. Bert's change in belief, in this case, seems to have been caused wholly by facts that are not reasons bearing on the change, such as the fact that a nefarious neuroscientist has stimulated such-and-such region of his brain.

D. Revising the Preliminary Sufficient Condition

My preliminary sufficient condition fares well when judged against the three cases introduced in Subsection II.A above. It does, however, fail to capture some cases of mental influence in which it seems clear that the mediating process is not at all sensitive to reasons. Consider the following case:

> *Brain Vibration*. Ada, a trustworthy source, arrives late to a meeting in a windowless room and says to all the attendees, as she enters the room, 'gosh, it's raining cats and dogs out there', as indeed it is. Bert doesn't hear this (his hearing is rather poor). In fact, he did not notice Ada entering the room at all. However, Ada's utterance nevertheless causes Bert to form the belief that it's raining. It does so in the following way: the precise combination of sound waves produced by Ada's utterance slightly vibrates Bert's brain and thereby stimulates some of his neurones to fire in just such a way as to produce the belief that it's raining.

Here, reasons for Bert's belief-change—namely the fact that it is raining and the fact that a trustworthy source has reported as much—causally contribute to the belief-change. Yet it is intuitively implausible that the mediating process within Bert is to any significant degree sensitive to those reasons or indeed other reasons. The mediating process in this case is the physical process via which the vibrations stimulate Bert's neurones, and this process seems to have produced an alteration that is supported by reasons through a kind of fluke, not through being sensitive to those reasons.

Consider also:

> *Reasons-Sensitive Neurostimulator.* All is as in the original *Neurostimulation* case, in which Cindy, a neuroscientist, produces Bert's belief that it is raining by electrically stimulating his brain, except in this case, Cindy is not indifferent to the evidence that it is raining. Rather, she only produces the belief that it is raining, in her subjects, when she recognises that there is good evidence that it is raining. On this occasion, Cindy produces the belief in Bert because she has just seen for herself that it is raining.

In this case, at least one reason that bears on the alteration—the fact that it is raining—causally contributes to the alteration of Bert's belief. However, it is once again implausible that the mediating process within Bert is sensitive

to the reason. Rather, it is Cindy, the neuroscientist, who is sensitive to the reason.

In order to yield the correct verdict on these cases, I suggest that we add the following supplemental clause to my preliminary sufficient condition: even when reasons do significantly causally contribute, via mediating process *P*, to the resulting alteration, *P* counts as reasons-insensitive if *P* was not a reasons-tracking type of process—that is, a type of process that is significantly and intrinsically disposed to translate reasons of the sort that were in fact causally operative into mental alterations that conform to those reasons. This leaves us with the following sufficient condition: mediating process *P* was reasons-insensitive if (i) reasons bearing on the resulting alteration did not significantly causally contribute, via *P*, to the alteration's occurrence, or (ii) *P* was not a reasons-tracking type of process.

Two clarifications are in order here. First, my revised sufficient condition employs the concept of an *intrinsic disposition*. What is an intrinsic disposition? I will adopt the standard answer to this question (Vihvelin 2013, p. 175): the intrinsic dispositions of an entity are its dispositions that are necessarily shared by other entities that have the same intrinsic properties and are governed by the same laws. The paradigmatic dispositions possessed by medium-sized options—such as the fragility of a crystal vase, or the flammability of a match—are intrinsic dispositions, and, indeed, it is controversial whether extrinsic dispositions exist (Vihvelin 2013, p. 175). A feature of intrinsic dispositions is that they are invariant to ordinary changes in circumstances. A crystal glass remains fragile (disposed to break) even if it is packed in layers of bubble wrap. A match remains flammable (disposed to ignite) even if it is in a tank from which all oxygen has been removed. Just as a vase can be intrinsically disposed to break, I assume that types of process, such as conscious reflection, reasoning, or perception, can be intrinsically disposed to translate reasons, or certain kinds of reasons, into mental alterations that conform to those reasons. Moreover, just as glass remains fragile even if it fails to break when dropped, a type of process remains intrinsically disposed to produce reasons-conformity even if one of its instances fails to do so because, say, there was a lot of misleading (apparent) evidence present. And just as something can break when dropped even if it is not significantly fragile—for example, because it was dropped from an extreme height or simultaneously struck by a meteorite—so too, a type of process can produce an alteration that is supported by reasons even though it was not significantly intrinsically disposed to do so; as I will explain below, I think this is precisely what happens in *Brain Vibration* and *Reasons-Sensitive Neurostimulator*.

The second clarification concerns the relationship between the two prongs—(i) and (ii)—of my sufficient condition. I have presented these prongs as distinct. There are two distinct ways in which a mediating process can qualify as reasons-insensitive: (i) reasons bearing on the alteration did not significantly causally contribute, via the mediating process, to the alteration's occurrence, or (ii) the process was not of a reasons-tracking type. However, it is possible to characterise the condition in a more unified way. The general idea behind the condition is that *P* is reasons-insensitive if it is not manifesting an intrinsic disposition of *P*-type processes to translate (certain) reasons into mental alterations that conform to those reasons. The two prongs of my condition capture two ways in which *P* could fail to manifest such a disposition: by not being a type of process that *possesses* the disposition to a significant degree (captured by prong (ii)), or by not *manifesting* that disposition through a significant causal contribution of reasons to the production of the alteration (captured by prong (i)). Nevertheless, in seeking to apply my revised sufficient condition, I will, for ease of exposition, generally keep prongs (i) and (ii) separate.

My revised sufficient condition—henceforth sometimes the 'causal-dispositional condition'—is satisfied in *Brain Vibration* and *Reason-Sensitive Neurostimulator*, for in neither of these cases is the mediating process of a reasons-tracking type. The sorts of physico-chemical processes involved in these cases *can* produce mental alterations that conform to at least some reasons—as they do in these cases—but when they do so, they are not manifesting any significant intrinsic disposition to do as much. In *Brain Vibration*, it is simply a fluke that the process produces an alteration that conforms to some reasons, and in *Reasons-Sensitive Neurostimulator*, it is due to the intrinsic disposition of Cindy, the neuroscientist, to produce reasons-conformity. Thus, the revised sufficient condition allows us to correctly classify these cases as cases in which the mediating process is reasons-insensitive.

I believe that my revised sufficient condition captures most, and perhaps all, cases of mental influence that we should want the insensitive process condition to capture—that is, cases in which it is intuitively clear that the mediating process is not significantly sensitive to reasons. However, I remain open to the possibility that it does not; as noted above, I recommend it as a sufficient, not as a necessary, condition for reasons-insensitivity. In the next subsection, I will consider one type of case that my condition does not capture, but that might be thought by some to involve a reasons-insensitive mediating process.

E. Sensitivity to a Narrow Range of Reasons

I have just suggested that my preliminary sufficient condition for reasons-insensitivity needs to be broadened to cover cases in which reasons significantly causally contribute, via mediating process *P*, to an alteration, but *P* was not significantly predisposed to produce reasons-conformity.

In this subsection, I consider the possibility that the condition should be broadened further still. There is a class of mediating processes that do not meet even my revised sufficient condition—the causal-dispositional condition—but that some might take to be 'reasons-insensitive'. I am thinking here of cases in which the mediating process is sensitive to only a narrow range of reasons.

Consider:

> *Salt Injection.* Franz, like all of us, possesses a negative feedback mechanism that controls his feelings of thirst. When his blood salt concentration is high, for example due to dehydration, sensor cells in his blood vessels detect this and send a hormonal signal to the brain which produces feelings of thirst. Suppose that Eliza causes Franz to feel thirsty by injecting salt into his blood without his consent. The high salt concentration she produces is sufficient to cause damage if it is not quickly reduced through fluid intake. Moreover, Franz is not likely to drink enough fluid quickly enough unless he experiences a strong feeling of thirst. Franz's sensor cells detect the increase in blood salt concentration and produce a strong feeling of thirst via the usual mechanism.

Given the background that the high salt concentration is sufficient to cause organ damage and that a strong feeling of thirst would help to prevent that damage, Franz's high blood salt concentration may constitute a reason for him to bring about or undergo an increase in his feelings of thirst (this could be disputed, but I find it plausible). Moreover, the neuroendocrine process that occurs within Franz is, it seems, significantly sensitive to this reason. The reason makes a significant, indeed very substantial, causal contribution to producing the alteration to Franz's feelings of thirst. Furthermore, the process was of a reasons-tracking type; such neuroendocrine processes have an intrinsic tendency to produce levels of thirst that are supported by the kinds of reasons in play in this case—reasons concerning blood salt concentration.

Yet it might be thought that, intuitively, the mediating process here is reasons-insensitive and moreover that Eliza has mentally interfered with

Franz. Significant reasons-sensitivity might be thought to require sensitivity to a broad range of reasons. But the neuroendocrine process that occurs within Franz is not broadly sensitive. It is highly sensitive to reasons that concern blood salt concentration.[9] However, it is insensitive to other reasons that plausibly bear on the alteration to Franz's thirst level. Perhaps Franz is about to begin a trek across a desert and needs to be as well hydrated as possible when he begins. This would plausibly constitute a reason for bringing about or undergoing the increase in his feelings of thirst. But it is doubtful that the neuroendocrine process that occurs within Franz is at all sensitive to this reason.

Should I, then, broaden my sufficient condition for reasons-insensitivity to capture cases like *Salt Injection*? I am not convinced that I should, since, as elsewhere in this book, I want to err on the side of caution, and I do not think it is *clear* that the mediating process within Franz is reasons-insensitive, nor do I think it is *clear* that there is mental interference in this case. I do think it is clear that Eliza *wrongs* Franz in *Salt Injection*. However, it is not obvious that the wrong is an infringement of the RAMI. (It may instead be only an infringement of the RABI.) Moreover, even if it *is* an infringement of the RAMI, it may not be one that is best captured in terms of the reasons-(in)sensitivity of the mediating process. Here, as elsewhere, I remain open to the possibility that there are RAMI-infringements besides those captured by my conditions.

III. Conclusion

In this chapter, I have further developed two conditions in my working account of the RAMI: the significance condition, which is one of three jointly sufficient conditions for a mental interference to infringe the RAMI, and the insensitive process condition, which is one of two jointly sufficient conditions for an action to qualify as a mental interference. In discussing the former condition, I identified a number of factors that may affect the magnitude of a mental interference, and thus whether it counts as significant: the number of mental states altered, the importance of the altered mental states, and the degree of the alteration. Turning to the insensitive process condition,

[9] This may somewhat understate the range of reasons to which the mediating process was responsive. Blood volume and blood salt concentration affect thirst responses via similar neuroendocrine pathways, so it may be that the process was also sensitive to reasons pertaining to blood volume. See, for a description of the neuroendocrine regulations of thirst in mammals, Todini and Fantuz (2023).

I offered a sufficient condition for a mediating process to count as reasons-insensitive, and thus for a mental influence to satisfy the insensitive process condition. I proposed that a mediating process is reasons-insensitive if either (i) reasons bearing on the alteration did not significantly causally contribute, via that process, to the occurrence of the alteration, or (ii) the process was not of a reasons-tracking type—that is, a type of process that is significantly and intrinsically disposed to translate reasons of the sort that were causally operative into mental alterations that conform to those reasons.

With these further details in place, I am now in a position to begin to draw out some of the concrete implications of the RAMI. That will be my primary task in the next chapter.

7
Implications of the Right Against Mental Interference

On the account that I have been developing, *A* mentally influences *B* if

> *A*'s action *X* alters a mental state of *B*'s (the *agential alteration condition*)

and this mental influence constitutes a mental *interference* if, in addition,

> the mediating process—the process within *B* via which *X* produces the alteration—is insensitive to the reasons that bear on the alteration (the *insensitive process condition*)

where it is sufficient for satisfaction of the insensitive process condition that

> either (i) reasons bearing on the alteration did not significantly causally contribute, via the mediating process, to the occurrence of the alteration, or (ii) the mediating process was not of a reasons-tracking type (the *causal-dispositional condition*).

Further, this mental interference infringes *B*'s right against mental interference (RAMI) if,

> *A* is a well-situated moral agent (the *duty-bearer condition*)[1]

[1] Recall that I am taking the class 'moral agents' to include at least all adult human beings who are not severely cognitively or morally impaired, and am assuming that these beings are also 'well-situated' if (i) they could reasonably be expected to foresee the consequences or other features of their conduct by virtue of which they interfere with the putative right-holder's mind in a particular case, and (ii) there is nothing severely impeding their exercise of their cognitive and moral capacities.

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0007

and

A's interference with *B*'s mind is significant (the *significance condition*)

and

B possesses the RAMI and has not waived or forfeited it in relation to the interference in question (the *nonliability condition*).

These conditions tell us, in abstract terms, when an action infringes the RAMI. But their implications in many particular cases are not obvious. Which interventions, concretely specified, meet these conditions, and which do not? This chapter will offer some preliminary answers to this question, focussing especially on cases in which I think the insensitive process condition will likely be decisive in determining whether the RAMI is infringed.[2]

In Section I, I offer some examples of mental influences that likely would *not* meet my jointly sufficient conditions for RAMI-infringement, since they are unlikely to satisfy the insensitive process condition. In these cases, the mediating process is likely to be significantly sensitive to the reasons that bear on the resulting alteration.

In Section II, I give some examples of mental influences that likely *would* satisfy the insensitive process condition, and thus to infringe the RAMI, provided that the duty-bearer, significance, and nonliability conditions are also satisfied, as I think they will be in many cases.

Both Sections I and II deal chiefly with types of influence on which I think my account will typically issue clear and plausible verdicts. In Section III, I turn to consider some harder cases—some types of mental influence on which my account may seem to offer either no clear verdict or a verdict that is not clearly plausible.

I. Mental Influences That Do Not Satisfy the Insensitive Process Condition

Let me begin, then, with some examples of mental influences that likely will not satisfy my jointly sufficient conditions for RAMI-infringement, since they likely

[2] I focus on this condition because I think it is the condition, among those listed above, whose concrete implications are least clear.

do not meet the insensitive process condition; the mediating processes involved in producing these alterations are likely to be significantly sensitive to reasons. I divide my examples into three groups, discussed respectively in Subsections I.A, I.B, and I.C. In suggesting that these influences likely operate via significantly reasons-sensitive mediating processes, I will rely heavily on intuition and anecdotal observation; however, in Subsection I.D, I will consider an objection to my claims that is grounded in the empirical cognitive science literature.

A. Explicitly Reporting Reasons

One way in which we can alter others' mental states is by explicitly reporting independent reasons for that alteration—reasons that would have obtained regardless of whether we had reported them. I have already given one example of an influence of this kind—the *Rain Claim* case from the previous chapter, in which Ada caused Bert to believe that it is raining by reporting as much. Other influences that take this form would include causing a person to believe that anthropogenic climate change is occurring by reporting that most climate scientists believe this, causing a person to intend to pick up the litter that they have just dropped by informing them that littering is legally prohibited, and causing a person to want to donate to a charity by pointing out the number of lives they could save by doing so.

In cases like these, at least two different reasons for the mental alteration are typically present: the fact being reported by the influencer, and the fact that a somewhat trustworthy source is reporting it. For example, in *Rain Claim*, the fact that it is raining is one reason for Bert to believe that it is raining, and the fact that Ada, a trustworthy source, has claimed that it is raining is another.[3] And in most such cases, these reasons will causally contribute to the alteration. In *Rain Claim*, for example, the fact that it is raining causally contributes to the trustworthy source—Ada—reporting that it is raining, which in turn causally contributes to Bert's coming to believe that it is raining. Thus, both the fact that it is raining and the fact that a trustworthy source has reported as much causally contribute to the alteration of Bert's belief (though the former fact may contribute only via the latter). Of course, *Rain Claim* is an unusual case, in that Bert engaged in conscious reflection on his reasons; we do not normally take such a careful and explicit approach

[3] I am not suggesting here that these reasons are independent of one another, nor that they should be treated as additive.

to updating our beliefs. In a more typical variant of *Rain Claim*, Bert's response to Ada's claim would be a rapid, largely subconscious one. Still, even in such a case, we would expect the fact that a trustworthy source has claimed that it is raining, and so indirectly the fact that it is raining, to make a causal contribution to Bert's belief-change. Moreover, even if Bert had responded to Ada's claim in a rapid, subconscious way, we would expect that Bert would at least implicitly weigh the reason that has been reported and/or the fact that it has been reported in the balance with other reasons for and against the alteration. We would, for example, expect his response to Ada's claim to be affected by the prediction from his weather app and the dark clouds that he had seen gathering; those facts would make him more inclined to form the belief that it is raining than he would have been in their absence.

I take it, then, that in cases of influence via explicitly reporting reasons, it is likely that a range of reasons will significantly causally contribute to the alteration via a kind of (often implicit) weighing process. Moreover, this weighing process will typically be of a reasons-tracking type; it will be significantly and intrinsically disposed to produce conformity with reasons of the sort that figure in the weighing.

B. Implicitly Drawing Attention to a Reason

Another way in which we commonly alter others' mental states is by implicitly drawing the influencee's attention to a reason for bringing about or undergoing the alteration. Again, I have already given an example of this:

> *Rain Gesture*. Ada arrives late to a meeting in a windowless room, drenched from the rain outside. As she enters the room, she gestures towards her soaked clothing with the intention of communicating to the attendees that it is raining outside. Bert, whose reliable weather app had reported a 93% chance of rain today, who had seen dark clouds gathering when he entered the building a short time earlier, and who knows that Ada is a trustworthy individual, sees Ada's gesture. Without any mediating conscious thought, he comes to believe that it is raining outside, whereas he did not previously believe this.

Here are some other examples: causing someone to experience a pang of guilt by nodding to the litter he has just dropped, causing someone to want to drive more slowly by placing a cross at the site of a fatal road collision, and causing someone to intend to donate to a charity by telling them a true

story in which a child's life was saved by someone else's donation to that same charity.

Again, in cases like these, there will often be at least two reasons for the influencee to undergo or bring about the alteration: the reason that the influencer draws attention to, and the fact that someone trustworthy is drawing attention to it. For example, telling someone a story about the life-saving effect of another's charitable donation draws the hearer's attention to a reason to intend to donate: that donating can have major benefits for the recipient of the donation. But the fact that someone trustworthy is drawing attention to this reason could be a further reason. The telling of this story plausibly constitutes an implicit recommendation, by the storyteller, to the story-hearer. And to the extent that the storyteller is trustworthy, we might think that the fact that the storyteller is recommending donation is a reason for the hearer to form the intention to donate.

In many such cases, at least one of these reasons will significantly causally contribute to the alteration. In many such cases, moreover, the influencee will implicitly weigh any reasons that have been presented in the balance with other reasons, so these other reasons will also causally contribute to the alteration. For example, suppose our story-hearer already knows that relieving poverty can have many knock-on benefits besides those which the storyteller has highlighted. This will likely incline her more strongly to forming the intention to donate, and thereby causally contribute to her formation of that intention. And in typical cases, the implicit weighing process engaged in by the influencee will be of a reasons-tracking type; processes like this don't just sometimes happen to produce mental alterations that conform to reasons of the sort that were operative here, they are intrinsically and significantly disposed to do so.

Of course, there will be cases in which the process mediating the alteration is not of a reasons-tracking type. Suppose, for example, that the national park authorities display a picture of a well-known mildly poisonous spider at the beginning of a forest path. They do this with the intention of drawing hikers' attention to the presence of these spiders in the forest and thereby inducing hikers to form an intention to be vigilant. But suppose that Jarvis, a hiker, is arachnophobic. The picture does induce in him an intention to be vigilant, but it does so by immediately triggering an overwhelming and difficult-to-alter fear of the mildly poisonous spiders. Here, reasons—such as the fact that there *are* mildly poisonous spiders in the forest, and the fact that the authorities recommend vigilance—causally contribute in a significant way to the hiker's formation of the intention. But the process via which the intention

is formed is not of a reasons-tracking type. It is rather a type of process that is disposed to produce exaggerated responses to reasons of the sort that were operative in this case.

There is, however, no reason to suppose that responses like that of the arachnophobic person are the norm. In typical cases where we influence others by drawing their attention to reasons, we do so via processes that are significantly and intrinsically disposed to translate (certain types of) reasons into conformity with those reasons.

C. Creating New Reasons

We can also alter others' mental states by creating new reasons in favour of the alteration—where by that I mean performing an action that creates a reason for the alteration, and does so other than by reporting or drawing attention to an independently present reason.[4] Examples would include causing someone to believe they are having a good time by telling a funny joke, causing a customer to intend to give you a tip by providing her with excellent service, and causing a person to prefer to leave their job by offering them a better one. As the last item on this list suggests, one way to create a new reason is to make available some desirable option. A common way to do this is to attach some reward or incentive to an already-available option. Thus, this third group of cases also includes standard cases in which one influences another through incentivisation, as when one causes a person to want to jump by offering him £10 if he can jump up and touch the ceiling.

When one influences another's mental states by creating a new reason, that new reason will typically causally contribute, in a significant way, to the alteration. Moreover, in most cases, other reasons that bear on the alteration will also causally contribute; if I already have many good reasons to quit my job, then your offering me a better job is more likely to result in me preferring to quit than if I lack these other reasons. The many other good reasons will play a role in producing my preference change. Why do these reasons causally contribute to the alteration? Typically because the influencee engages in at least a rough-and-ready implicit weighing of reasons—a type of process

[4] When a trustworthy source reports or draws attention to some pre-existing reason, the fact that *p*, they are, as I have already noted, typically also creating a reason—the reason being the fact that someone trustworthy has reported or drawn attention to *p*. However, I do not count such cases as instances of creating a *new* reason, since the primary reason operative in such cases is one that would in any case have been present.

that is significantly and intrinsically disposed to translate reasons into conformity with those reasons. Thus, in typical cases, my sufficient condition for reasons-insensitivity will not be satisfied. Reasons will make a significant causal contribution, via a weighing process, to the occurrence of the alteration, and the weighing process will be of a reasons-tracking type.

Three special cases of mental influence that operate via creating new reasons warrant specific mention, since it is perhaps not obvious that they fall within this category. The first case is the case of inducing a person to possess a desire, intention, or other motivation to do something by requesting or demanding that they do that thing. Suppose that your employer causes you to want to stay at the office by saying 'Could you please stay a little longer and finish off this report?' or simply 'Stay and finish the report'.[5] At least in certain contexts—especially those in which the speaker has some form of authority—utterances such as these create reasons for hearer to alter their motivational states, and I take it that in typical cases of this kind, the mediating process will be sensitive to those reasons.

A second special case is the case of causing someone to believe something by saying something that makes it true, or likely to be true, as, for example, where one member of a couple causes the other to believe that they are married by saying 'I do', or where a mayor causes a citizen to believe that the park is called 'The Green' by saying 'I hereby name this park "The Green"'.[6] Again, I take it that, where such utterances indeed create reasons, the mediating response in the hearer is likely to be sensitive to the reason(s) created.

The third special case that warrants mention is the case of inducing a perceptual state through the usual perceptual processes. Suppose I cause someone to perceive that there is a red flag in front of her by waving a red flag in front of her, which she perceives in the usual way. I think this case is again just a special instance of the type of influence that I have been describing in this section: influencing by creating new reasons. When I wave the red flag in front of a person's face, I create a reason for the person to perceive that there is a red flag in front of her—the reason being the fact that there *is* a red flag in front of her.[7] This reason, it seems, makes a substantial causal contribution to

[5] Many cases of this kind, including the examples I mention here, involve what J. R. Searle called 'directives': speech acts that attempt to get the hearer to do something. See, for the classic discussions, Searle (1969, 1976).

[6] Many cases of this kind, including the examples I mention here, involve Austinian performatives; utterances that act on, and do not merely describe, the world. See, for the classic discussion, Austin (1975).

[7] There may also be other reasons created here. For example, there may be instrumental reasons for the person to perceive the flag since, if she does not, she might walk into it, or fail to appreciate some message that the flag waver is trying to send.

the alteration in her perceptual states. And moreover, the perceptual process that produces the alteration is of a reasons-tracking type; perceptual processes do not involve anything that could aptly be characterised as 'weighing reasons', but they do still tend to translate reasons of the kind that are causally operative in producing perceptual states—namely, facts about the environment that we are perceiving—into mental alterations that conform to those reasons.

It is true that some take perception to be a reasons-insensitive process. However, this, I think, is because they are operating a more inclusive understanding of reasons-insensitivity than I am.

One basis for taking perception to be a reasons-insensitive process would be that it involves no conscious reflection on one's reasons. The thought could be that significant reasons-sensitivity requires such reflection. I allow that a reflection-requiring conception of (significant) reasons-sensitivity might be apt for some purposes. However, I think it is clear that, whatever its merits in other applications, it is not the conception that is relevant for establishing infringements of the RAMI. This is because it is clear that some mental influences operate via processes that involve no conscious reflection on reasons without thereby qualifying as mental interference. It is clear, for example, that Ada's intervention in *Rain Gesture* is not an instance of mental interference. And this is clear even though the mediating process does not involve conscious reflection on reasons.

Another basis for taking perception to be a reasons-insensitive process would be that it is sensitive only to a narrow range of reasons. Perceptual processes are sensitive to perceptual stimuli but not to other types of reason. For example, they are not clearly sensitive to higher order evidence. Sometimes, as when presented with a known optical illusion, we have evidence that our perceptual processes are not tracking reasons, yet our grasping this evidence often does not, or not immediately, alter our perceptual experiences.[8]

I concede that perceptual processes are somewhat limited in the range of reasons to which they are sensitive. However, as I explained when discussing the *Salt Injection* case in Chapter 6, Subsection II.E, my causal-dispositional condition for reasons-insensitivity does not classify a process as reasons-insensitive merely because the range of reasons to which it is sensitive is

[8] It is often claimed, more generally, that perceptual processes are 'encapsulated'—walled off from 'central cognition', where the domain-general assessment of reasons occurs. Whether perceptual processes are indeed encapsulated, and, if so, how far this restricts their reasons-sensitivity is a matter of dispute. For a helpful discussion, see Jenkin (2023).

limited. And I do not think that it should, since I don't think it is clearly apt to characterise such processes as reasons-insensitive, nor do I think that mental influences which operate via such processes clearly qualify as mental interferences.

I will assume, then, that normal perceptual processes do not typically satisfy my insensitive process condition, so mental influences that employ them do not typically qualify as mental interferences.

Let me end this subsection by commenting on a type of influence that operates via creating new reasons, yet might be thought to involve mental interference: coercion. Consider the oft-invoked highwayman case: a robber holds you at gunpoint and says 'your money or your life'. Suppose the robber thereby succeeds in altering your intentions—from not intending to hand over your wallet, to intending to do so. The robber exerts a mental influence on you, and some might think that this mental influence amounts to a mental interference (I don't myself have a clear intuition on this). Yet the robber influences your mind through creating a (very strong) reason in favour of the alteration to your intentions, and that reason makes a major causal contribution to the alteration. Is the process via which the threat alters your intentions of a reasons-tracking kind? That depends. It could be that it is not; perhaps the threat will induce a state of panic in which various random thoughts pop into your head, including the intention to hand over your wallet. But it is also quite possible that the process will be of a reasons-tracking kind; perhaps you will relatively coolly reason to the conclusion that passing over your wallet is the best course of action available to you. In that case, the insensitive process condition will not be met, and my account will not classify your threat as a mental interference. I am happy with that result, since I do not myself find it clear that such cases should be classed as mental interferences. However, if others are less happy, that need not be a problem for my account. Recall that my account is intended only as offering a set of jointly sufficient conditions for mental interference. Thus, it is compatible with my view that there are other forms of mental interference, not captured by my account. Mental influence via coercion could perhaps be one of these.

D. An Objection from Cognitive Science

In suggesting that the three types of mental influence distinguished above would be unlikely to meet the insensitive process condition, I have relied heavily on intuition and anecdotal observation. In this section, I wish to

consider an objection to my claims that is motivated by empirical findings from the cognitive sciences. These findings suggest that, even the mental influences that seem most likely to operate via significantly reasons-sensitive processes—instances of altering other's beliefs by explicitly reporting reasons—may exert their mental effects via mediating processes that are less sensitive to reasons than is commonly assumed.

In the cognitive science literature on how we update our beliefs in response to declarative assertions by others—like Ada's assertion that it is raining in *Rain Claim*—there is disagreement about whether what I above referred to as a 'weighing' process occurs primarily before the belief-change occurs (in determining whether to change one's beliefs) or primarily after the belief-change has already been tentatively made (in determining whether to reverse the change). These are often referred to as respectively the Cartesian and Spinozan models of belief-change. On the Spinozan model, which has significant support from empirical work in the cognitive sciences (Gilbert 1991; Mandelbaum 2014), the process of initial belief-change is an effortless and passive one, with the effortful weighing of reasons occurring only after the belief-change has occurred, in deciding whether to reject the new belief. Taking up the belief that is supported by the declarative assertion is the default.

If the Spinozans are correct, this might be thought to cast doubt on the idea that, in typical cases of mental influence via reporting reasons, the process producing the initial belief-change is significantly sensitive to reasons. The Spinozan would, for example, claim that, in an empirically realistic version of *Rain Claim*, Bert would, on hearing Ada's assertion, passively and effortlessly come to believe that it is raining, and would only then, in assessing whether to retain or reject this belief, engage in an effortful weighing of reasons. More generally, the Spinozan might claim that in typical cases of altering a person's beliefs by reporting reasons, the initial alteration will be produced via passive, effortless processes that are not significantly sensitive to reasons, with the perhaps implausible implication that the influence will qualify as a mental interference, on my account thereof.

I have two responses to this challenge. The first involves distinguishing the alteration that occurs before the new belief is assessed for rejection from the alteration that obtains afterwards. Suppose that, at t_1, you say 'it's hot outside', and I then passively and effortlessly, by t_2, form the belief that it's hot outside, before actively and effortfully assessing this belief for rejection, a process that is complete by t_3 and results in my retaining the belief. Even if the passive and effortless process via which my belief at t_2 is produced is reasons-insensitive,

the process via which my belief at t_3 is produced is not. Thus, though the influence on my t_2 beliefs may count as a mental interference, on my view, the influence on my t_3 beliefs does not. This is important since it suggests that the interference with my beliefs in this case affects only a limited time slice of my belief—the belief that obtains between t_2 and t_3—and this time slice will likely be short in duration. At worst, then, it seems that I am committed to saying that there is a short-lived, and therefore possibly insignificant, mental interference in this case.

Of course, there may be actual cases in which the process of assessing a belief for rejection never occurs; *A* reports a reason, *B* passively and effortlessly forms a belief as a result, and this belief never gets assessed for rejection. There is, for instance, some empirical evidence that the process of assessing a belief for rejection often does not occur when the believer is under cognitive load (see, e.g. Gilbert et al. 1990).[9] In such cases, the response I have just offered will not succeed. However, in cases where the assessment process *does* occur, I think it can plausibly be maintained that any mental interference will be short-lived and perhaps insignificant.

In fact, however—and here is my second response—I doubt that there is even an insignificant mental interference in such cases, since I doubt that the process via which beliefs are initially updated, on the Spinozan picture, is normally reasons-insensitive. Even if the process via which a belief is initially acquired is passive and non-effortful, it could still be significantly sensitive to reasons. There are, after all, other passive and non-effortful processes that are significantly sensitive to reasons (perception, if my comments in the previous subsection were on target, is one of these). Moreover, there is, I think, some positive basis for supposing that the initial updating process posited by the Spinozan is sensitive to reasons. The process via which we acquire beliefs in response to declarative assertions does, after all, appear to be somewhat *context*-sensitive, and many of the features of the context to which it is sensitive look very much like reasons. Suppose that you say 'it's hot' and I, as the Spinozan predicts, passively and effortlessly form a belief. *Which* belief I form will very much depend on the context. If you've just put a piece of quiche on my plate, I'll form a belief along the lines of 'the piece of quiche is hot'. If you've just come in from outside and wiped the sweat from your brow, I'll form a belief along the lines of 'it's a hot day out'. The belief I will form seems to depend on the other pieces of evidence available to me,

[9] I thank Henry Schiller for pressing me to consider this evidence. For a discussion of it, see Mandelbaum (2014, pp. 65–66).

suggesting that the process via which I form it, passive and effortless as it may be, is somewhat sensitive to those pieces of evidence—those *reasons*. This is of course consistent with its being the case that the subsequent assessment of whether to reject the belief is where sensitivity to *most* reasons enters the scene; this later assessment may be more sensitive to reasons, or sensitive to a wider range of reasons, than the process leading to the initial alteration in belief. But that initial process is still, I think, reasons-sensitive to a significant degree.

II. Mental Influences That Satisfy the Insensitive Process Condition

Having outlined some examples of mental influence that likely do *not* count as mental interferences or meet my conditions for RAMI-infringement, since they typically fail to meet the insensitive process condition, I now turn to consider some types of influence that I think typically do satisfy this condition, since the mediating process is not, in most cases, significantly sensitive to reasons. These mental influences thus typically count as mental *interferences*. Where they are also significant, and exerted by well-situated individuals on nonliable individuals, they meet my conditions for infringement of the RAMI.

I draw here chiefly on cases that I have offered in previous chapters.

Consider first some examples given in Chapter 1. There, I mentioned, as candidate infringements of the RAMI, the use of *nonconsensual neurointerventions*. As an example, I gave the administration of psychopharmaceuticals under mental health legislation to individuals deemed to pose a risk to themselves or others (we might think, for instance, of the administration of anti-psychotic medications to suppress persecutory delusions that are causing the deluded individual to regularly attack his loved ones). Another example would be the chemical castration of sex offenders in order to suppress their sex drive, in (partial) fulfilment of a criminal sentence.

In these cases, it is plausible that there are reasons for the influencee to bring about or undergo the resulting mental alteration. A person experiencing persecutory delusions causing him to regularly attack his loved ones plausibly has reasons to bring it about that he no longer experiences those delusions, and a sex offender may have reasons to bring it about that he experiences weaker sexual desires. However, it is doubtful, in these cases, that

the mediating process—the physico-chemical process within the influencee via which the neurointervention produces the alteration—is significantly sensitive to these and other reasons that bear on the alteration.

Applying the causal-dispositional condition for reasons-insensitivity introduced in the previous chapter, there are two questions to distinguish here. First, do reasons causally contribute, via the mediating process, to the alteration in these cases? And second, is that mediating process of a reasons-tracking kind? If the answer to either question is 'no', then these influences will come out as mental interferences, on my view.

Depending on how these cases are further spelled out, the answer to the first question may be 'yes'; reasons may causally contribute to the alteration in these cases. For example, the fact that the persecutory delusions cause a person to harm others—which is a reason to extinguish the delusions—may cause the doctors to administer the anti-psychotics that indeed extinguish them. However, the answer to the second question will, I think, typically be 'no'. The physico-chemical processes via which neurointerventions operate are not typically disposed to translate reasons into reasons-conforming mental alterations. Processes of these types *can* produce mental alterations that conform to reasons, but when they do, this is typically because the individuals administering the interventions are disposed to produce mental alterations that conform to reasons. The mediating processes themselves are not so-disposed.

Similar thoughts apply to many of the examples I gave in Chapter 4. Recall *Brain Drug*, in which your friend Lucien became gloomy following a period of marital tension, and you mitigated his gloom by spraying a fast-acting anti-depressant drug onto his skin. In this case, it is plausible that Lucien has reasons to be less gloomy and that these play a substantial role in causing you to intervene and thus in causing him to become less gloomy. However, the physico-chemical process via which your drug operates is not of a reasons-tracking type. It is you, not the mediating process within Lucien, that is sensitive to reasons here.

Finally, consider a case similar to the *Flashing Lights* case that I gave in Chapter 2. Suppose that I want a person to experience the mental components of an epileptic seizure, and I bring it about that she indeed experiences these by exposing her to flashing lights. The process via which this mental alteration is produced is unlikely to be sensitive to any reasons bearing on it. Thus, suppose that my victim happened to be driving a car at the time of my intervention, giving her an especially strong reason *not* to experience the seizure. It is highly doubtful that this will have made any negative causal

contribution to the occurrence of the seizure, for example, by making it less likely to occur. More generally, I see no basis for thinking that reasons bearing on the alteration in this case would have played any causal role in the neural processes via which the perception of flashing lights induces seizures. Thus, this mental influence will likely satisfy the insensitive process condition, and so count as a mental interference.

I have been suggesting, regarding the types of intervention discussed in this section, that the mediating process would 'typically' or 'likely' come out as insensitive to reasons, and thus that the interventions would 'typically' or 'likely' qualify as mental interferences, under my account. However, it is possible to imagine interventions of the kinds I have been discussing in this section—including neurointerventions—in which the mediating process is significantly sensitive to reasons. (In what follows, I will refer specifically to neurointerventions, but my comments would also apply, for instance, to the variant of *Flashing Lights* described above.)

Two types of case warrant mention. First, there are cases in which a neurointervention produces a mental alteration via the mediation of a significantly reasons-sensitive mental process such as reasoning or conscious reflection on one's reasons. Thus, suppose that chemical castration is administered to a sex offender with the goal of freeing the offender from overpowering sexual urges, thereby allowing him to form sex-related intentions on the basis of a calmer and less compulsive thought process. Suppose that it indeed has this effect: once the powerful sexual urges have been attenuated, the individual is able to think more clearly about his sexual conduct. And suppose that, as a result of this clearer thinking, the individual forms the intention to abstain from further sexual offending.

An intervention of this type will clearly produce multiple different mental alterations, but let us distinguish two: the initial effect on the person's sexual urges (i.e. the attenuation of the powerful urges), and the subsequent effect on the person's intentions (the formation of the intention to abstain from further sexual offending).

With regard to each of these alterations, we can ask: was the process that produced the alteration significantly sensitive to reasons? With respect to the second alteration, it seems clear that the answer is 'yes'. After all, the process via which the chemical castration produces this alteration includes the process of clear thinking that immediately led to the change in intentions. Thus, the mental influence exerted on this person's intentions with respect to sexual offending is unlikely to meet the insensitive process condition. However, the upstream mental influence on the person's sexual urges likely *does* meet this

condition. *This* alteration is brought about via a brute neurochemical—not mental—process that is likely not significantly sensitive to reasons.

Second, there are cases in which an intervention operates by creating a bodily state that is itself a reason that bears on the resulting mental alteration, and also causally contributes to it. Consider *Salt Injection*—a case discussed in the previous chapter. Recall that in that case, Eliza caused Franz to feel thirsty by injecting salt into his blood, thereby initiating a neuroendocrine process which produced feelings of thirst. Having a high level of salt in one's blood plausibly constitutes a reason to feel thirsty, at least where feeling thirsty helps one to drink enough to reduce the salt concentration to safe levels, and in this case that reason appears to causally contribute to the production of a feeling of thirst. Moreover, the neuroendocrine process that produces this alteration was significantly and intrinsically disposed to translate reasons of the sort that are operative here—reasons pertaining to blood salt concentration—into mental alterations that conform to those reasons. Thus, cases like this will, as I have already indicated, not satisfy my insensitive process condition.

III. Some Harder Cases

The influences discussed in the previous two sections were, for the most part, relatively easy cases for my account of mental interference. For most of these influences, my account yielded relatively clear verdicts on whether the influence in question is likely to count as a mental interference under the account. Moreover, these verdicts were, I hope, more-or-less plausible.

In this section, I turn to consider some harder cases; some cases in which it is difficult to say what verdict my account would give, and/or in which it is unclear whether the verdicts yielded by my account are intuitively plausible. In the next chapter, I consider some cases in which my account may have implications that are frankly unacceptable.

A. Salience-Based Nudges

Consider first what we might call *salience-based nudges*: interventions that—when successful—cause a person to take an option by making it more salient to the person but without thereby significantly changing the costs or benefits associated with that option or the alternatives to it. I take making an option

more salient to a person to consist in causing the option, or some feature of it, to occupy a greater share of the person's attention.[10] A classic example of a salience-based nudge is what I will call *Cafeteria Nudge*: inducing some people to take healthier foods in a cafeteria by placing those foods in a prominent location—for example, at eye level.

In the normal course of events, salience-based nudges have mental effects: they affect the attention that the nudgee gives to the various available options, and, when successful, they affect the nudgee's decisions and resulting intentions. However, they do not always clearly *alter* a person's mental states, as I understood 'altering' in Chapters 2 and 3. This is because they do not always cause the nudgee to have a mental state that clearly differs from the state that would have obtained in the baseline situation. (Recall that, in Chapter 2, I specified that mental alterations consist in bringing it about that a person has a mental state that differs from that in the baseline situation(s), while remaining neutral on how, precisely, we are to characterise that baseline.) In *Cafeteria Nudge*, for example, it is not clear that the situation in which healthy foods are at eye level is a deviation from the baseline; foods have to be arranged in one way or other, and the situation in which healthy foods are placed at eye level is as 'normal' and 'expected' as any other, so has a good claim to being considered a member of the set of baseline situations.

Salience-based nudges will produce a mental alteration, on my account thereof, only where the environment created by the nudge is a deviation from the baseline situation(s). Here is an example in which that is clearly the case:

> *Enhanced Cafeteria Nudge.* In order to encourage healthier food choices, cafeteria staff surround the green salads, but no other food items, with festive lights. The resulting lighting arrangement is significantly different from any plausible alternative—that is, any lighting arrangement that one might have expected to have been employed in the absence of the nudge. It also results in healthy foods appearing more salient than they would have appeared under any plausible alternative lighting arrangement. The nudge results in some customers choosing a green salad when they would not have done so under any plausible alternative lighting arrangement.

[10] I suspect that some of the considerations that I invoke in this subsection would also apply to other salience-based mental influences, for example, to linguistic devices which exert a mental effect by drawing something to the hearer's attention ('have you considered the weather?' or 'that apple might be rotten'). However, I will neither explicitly address such influences nor stake myself to any stance on them. For recent discussion of the rationality of our responses to some linguistic devices for inducing salience, see Dever and Schiller (2021).

It is clear that the situation produced by this nudge is not the baseline situation, nor among the set of baseline situations; in the baseline situation(s), either no festive lights would be used or at least they would be distributed more evenly. Moreover, some customers make different choices when exposed to the nudge than they would have made in the baseline situation(s). Thus, in this case, cafeteria staff do clearly *alter* the mental states of at least some customers, and so their action clearly satisfies the agential alteration condition and thus counts as a *mental influence*, as I have been using that term. Does it also satisfy the insensitive process condition, and so count as a mental *interference*?

To answer this question, we first need to distinguish between a range of different mental alterations that the nudger produces in this case. Let us focus on a single cafeteria customer who would not, in the baseline situation, have decided to take the green salad, but who does, in the presence of the nudge, make that decision. In relation to such a customer, the nudger plausibly produces all of the following mental alterations:

- The nudgee sees that the green salads are surrounded by festive lights when she would not have seen this in the baseline situation.
- The nudgee pays more attention to the green salads than she would have paid them in the baseline situation.
- The nudgee is more motivated to take a green salad than she would have been in the baseline situation.
- The nudgee forms the intention to take the green salad when she would not have formed this intention in the baseline situation.

These different mental influences may differ with respect to the reasons-sensitivity of the mediating process.

Consider, to begin with, the mental influence that consists in inducing the nudgee to see that the green salads are surrounded by festive lights. This is just a straightforward case of the type of influence that I discussed in Subsection I.C above: inducing a perceptual state by presenting a perceptual stimulus. I suggested above that such influences will not typically meet the insensitive process condition. The nudgee has a reason to see that the salads are surrounded by festive lights—the reason being that they indeed are—and the nudgee is likely significantly sensitive to that reason.

However, things may be different in relation to the mental influence that consists in inducing the nudgee to form the intention to take the green salad.

It is not so clear that the process leading to the formation of that intention is significantly sensitive to the reasons bearing on *that* alteration.

Let us consider two types of reasons that bear on this alteration, and to which the mediating process within the nudgee may be sensitive.

First, there are pre-existing reasons bearing on the alteration. These may include pre-existing reasons to intend to take the green salad (such as the fact that the green salad is healthier than most other options), and pre-existing countervailing reasons (such as the fact that one is allergic to celery, which the salad contains). Is the mediating process that occurs within the nudgee sensitive to these pre-existing reasons? It might be, but this will depend on the details—it will depend on how, exactly, the salience of the green salad leads to the nudgee forming the intention to take one. There are certainly ways of spelling out this process such that it will clearly be significantly sensitive to pre-existing reasons. One possibility, for example, is that the nudge works by triggering something like reflection on one's reasons. Perhaps the salience of the green salads prompts the nudgee to consciously reflect on whether she ought to take a green salad, and perhaps the intention to take one is simply the upshot of that conscious reflection. If the mediating process involves conscious reflection in his way, it will likely be highly sensitive to her pre-existing reasons. On the other hand, if the mediating process is an entirely subconscious, automatic one, it is less clear that the process will be significantly sensitive to pre-existing reasons. However, it is still possible. The degree to which the processes mediating salience-based nudges are sensitive to pre-existing reasons is ultimately an empirical question, and it is one on which there is limited empirical evidence. Nevertheless, what evidence there is suggests that the processes mediating salience-based nudges typically are significantly sensitive to at least some kinds of reasons: there is evidence that these processes are typically sensitive to the pre-existing *preferences* of the nudgee, suggesting that they would be sensitive to the nudgee's *preference-based reasons* (De Ridder et al. 2022, pp. 350–51). To give but one example, Taube and Vetter (2019) found that presenting 'green' products as the default option had a stronger effect on individuals with pre-existing pro-environment attitudes than on others; that is, it was more likely to be effective in inducing selection of 'green' products in those with the pre-existing pro-environment attitudes. Conversely, presenting non-'green' products as the default had a weaker effect on that same group. The nudgees' environmental preferences seem to have made a difference to the nudge's effects.

Neil Levy (2019) has suggested that there may be a second type of reason in play in many nudges. He argues that nudges are often, and are often

perceived by nudgees to be, implicit *recommendations*, and there is some empirical support for the view that they are indeed often perceived as such.[11] In relation to *Enhanced Cafeteria Nudge*, the thought would be that, in setting up the festive lighting so as to make the green salads especially salient, cafeteria staff are implicitly recommending those salads. And the fact that the nudger is recommending the green salads plausibly counts as a reason for the nudgee to decide to take one—at least in cases where the recommender is trustworthy. On my causal-dispositional condition, whether the nudgee is sensitive or insensitive to this reason would depend, in part, on whether the fact that the green salad is being recommended makes a causal contribution to the nudgee's mental change. If, for many individuals, the presence of the recommendation does make a causal contribution, then we might expect that the effectiveness of the nudge could be enhanced by making it clearer to nudgees that something is being recommended (e.g. by making it clear that the lighting has been arranged to promote healthier choices) and conversely that the effectiveness of a nudge could be decreased by presenting nudgees with (apparent) evidence that nothing is being recommended (e.g. by claiming that the lights have been arranged randomly). There is some empirical work on whether nudge-effectiveness is affected by making the intention—and thus recommendation—behind the nudge explicit, and thus has generally found that it is not. For example, Kroese et al. (2016) found that an experimental food-salience nudge similar to *Cafeteria Nudge* was effective, and equally so, regardless of whether the experimenters' health-based reasons for arranging the food as they did were disclosed. Similar results have been found for other salience-based nudges.[12] This could be taken to suggest that nudgees are not sensitive to whether the salient item is being recommended (though it could also be that the nudgees are so likely to assume that it is being recommended, even when the recommendation is left implicit, that making the recommendation explicit makes no appreciable difference). On the other hand, McKenzie and collaborators (2006) uncover some self-report-based evidence that perception of a recommendation does make a causal contribution to a nudge's effect. Overall, I take the empirical evidence to be inconclusive on whether the processes that mediate salience-based nudges are typically sensitive to the implicit recommendation that is present in these nudges. However, since evidence does suggest that they are typically sensitive to at least some pre-existing reasons—the nudgee's

[11] See, for example, McKenzie et al. (2006).
[12] See, for example, Bruns et al. (2018) and Steffel et al. (2016).

preference-based reasons—I think that salience-based nudges will not normally satisfy the insensitive process condition.

B. Aversion Therapy

The second type of influence that I wish to consider is aversion therapy, which I will take to consist in intentionally altering a person's mental states through a conditioning process that involves associating a particular type of mental state with another mental state that is aversive (i.e. negatively valenced), such as pain or disgust.[13] As with nudges, we need to distinguish different mental influences involved in aversion therapy. The two most obvious influences are (i) producing the aversive mental state (e.g. pain, disgust) and (ii) attenuating the mental states that are being 'conditioned away'.

Consider first the influence that consists in producing the aversive mental state. Whether this influence satisfies the insensitive process condition will depend on which aversive state is produced and how. If, say, the aversive state is disgust, and it is produced by presenting disgusting images, the mediating process will plausibly be significantly reasons-sensitive (that an image is disgusting plausibly counts as a reason for a person viewing the image to feel disgusted, and the process via which the influencee arrives at this feeling of disgust is plausibly significantly sensitive to this reason). On the other hand, if the aversive state is produced simply by electrically stimulating the pain or disgust centres in the brain, say, there is no reason to think that the mediating process will be reasons-sensitive.

Consider next the mental influence that consists in attenuating the mental states that are being conditioned away. To fix matters, let us focus on a case in which aversion therapy is used to treat problematic drinking in someone called Pim. Suppose that a desire for alcohol is elicited in Pim by exposing them to alcohol advertisements, and that, immediately thereafter, feelings of disgust are induced by exposing them to disgusting images. Moreover, suppose that, following many such exposures, Pim begins to experience weaker desires for alcohol when presented with the advertisements.

[13] Aversion therapy could be done with or without consent, and whether valid consent is present might be crucial in determining whether it infringes the RAMI. However, it is not, on my view, relevant to whether it constitutes mental interference, which is my chief focus here. Thus, I will not, in this subsection, pay much attention to whether the aversion therapy is consensual.

Our focus here is on the process within Pim via which Pim's desires for alcohol are attenuated. Is this process significantly reasons-sensitive? I think this will depend on the details of the process. One possibility is that Pim treats the aversive stimuli as a *disincentive* to persisting in their desire to drink. The repeated association of disgusting images with the desire to drink gives Pim reason to expect that future desires of this sort will be associated with further aversive stimuli, which arguably gives them a reason not to desire alcohol. If the mediating process within Pim involves his (explicitly or implicitly) recognising this reason, the process will be significantly sensitive to reasons.

However, it is a common feature of aversion therapy that it continues to exert its effect even once it is clear that the aversive stimuli are no longer going to be applied.[14] The effect that may initially be a rational response to disincentives subsequently becomes 'fixed'.[15] Suppose that this is true in our problem-drinking case: Pim's desires to drink remain suppressed even when it is clear that such desires will no longer result in exposure to disgusting images. That the effect on the individual's desires persists in this way suggests that, even if the early mental alterations produced by aversion therapy are produced via a process that is highly sensitive to at least one reason that bears on the alteration—the fact that it can reasonably be expected that future desires will be accompanied by aversive stimuli—the later mental alterations are not sensitive to that reason.

Thus, suppose that our programme of aversion therapy for Pim is administered from t_0 to t_1 and that from t_2 onwards, it is no longer reasonable for Pim to expect that future instances of the desire to drink will be accompanied by aversive stimuli. Consider now the alteration that the aversion therapy programme causes to Pim's post-t_2 desires—to the fact that Pim has a weaker desire to drink, after t_2, than they would have had in the baseline situation. *This* alteration appears to have been caused primarily by the fact that there *was previously*, prior to t_1, a correlation between the desire and aversive stimuli, but *that* fact is not a reason in favour of the post-t_2 desire-alteration. In relation to the post-t_2 desire-alteration, this fact about the past is simply a non-reason cause. Moreover, there is little reason to think that

[14] This phenomenon is sometimes characterised as 'behavioural momentum'. For a recent review of the literature on behavioural momentum following conditioning, see Trump et al. (2021).

[15] Some authors have, however, questioned this. Nick Chater (2009) marshals a range of experimental evidence against the orthodox view that conditioned responses are typically strongly persistent. He argues that most conditioned responses in humans can be quickly extinguished, with strong persistence being the exception rather than the rule.

the reinforcement learning or conditioning process via which this post-t_2-desire has been produced would be sensitive to other reasons bearing on the alteration.[16] Perhaps Pim's drinking is causing him problems in his relationships, giving him strong reasons to attenuate his desire to drink, but it is doubtful that the conditioning process would be sensitive to this reason. Indeed, conditioning processes are characteristically assumed to be paradigmatically reasons-insensitive processes.[17] Thus, it is quite possible that the influence on Pim's desires—at least, their post-t_2 desires—will come out as a mental interference on my view. And if it does, it will plausibly also qualify as a *significant* interference. Thus, if the aversion therapy is administered by a well-situated moral agent (so that the duty-bearer condition is satisfied) and Pim possesses the RAMI and has not waived or forfeited it in respect of the aversion therapy (so that the nonliability condition is satisfied), there would be an infringement of the RAMI in such a case.

It might be thought that the practical significance of this conclusion is limited, since the nonconsensual use of aversion therapy is rare. However, there are some rather common forms of influence that are posited to operate via similar mechanisms—that is, forms of conditioning—and in some cases, it is doubtful that valid consent has been obtained for these. Consider, for instance, the use of randomised rewards, often in the form of 'loot boxes', in video games.[18] Many video game designers randomise the reward (e.g. the number of points, tokens, or weapons) that a player receives for realising some achievement (e.g. passing a level[19] or accumulating some level of in-game currency[20]) in the game. Randomised rewards are often employed with the aim of increasing the amount of time that a player will spend playing a

[16] I henceforth use only the term 'conditioning', but readers who take there to be a distinction between conditioning and reinforcement learning can read all references to conditioning as references to conditioning and/or reinforcement learning.

[17] See, for example, Quilty-Dunn and Mandelbaum (2018, §2.2). Some types of reason do, however, seem capable of quickly extinguishing many conditioned responses. For example, as Nick Chater (2009, p. 355) notes, a slot machine user presented with the message 'out of cash, call supervisor', will immediately cease feeding money into the machine. This suggests a degree of reasons-sensitivity. Chater argues that, as a general matter, psychologists have underestimated the reasons-sensitivity of conditioning processes.

[18] I focus on the use of randomised rewards in video games. However, similar techniques are also used in other contexts—such as board games (consider, e.g. chance cards in Monopoly) and gambling—and much of my analysis would apply to those uses, too. For a brief survey of the historical uses of randomised rewards, see Nielsen and Grabarczyk (2019, pp. 3–4).

[19] This is one way in which loot boxes can be acquired in the game Overwatch. See Lawrence (2017) and Wiltshire (2017).

[20] This is one way of accessing loot boxes in the game Hearthstone (though there is also an option to buy 'card packs' with real-world money). See Wiltshire (2017).

game, since there is evidence that randomising rewards creates a more persistent motivation to continue the activity being rewarded (Ferster & Skinner 1957).[21] The greater persistence of the motivation is normally assumed to be the result of a conditioning process,[22] suggesting that its induction through the use of randomised rewards might—like aversion therapy—meet my insensitive process condition, and thus qualify as a mental interference. Moreover, it is doubtful that all gamers have validly consented to being subject to this interference. It might be claimed that merely choosing to play a game constitutes valid implicit consent to the use of randomised rewards in settings where (i) the player knows (or could reasonably be expected to know) that game designers often employ randomised rewards and (ii) the player knows (or could be reasonably expected to know) roughly what effects those rewards have on the motivation to continue playing the game. It is, however, very plausible that many players fail to satisfy at least one of these conditions, and for these players, it is doubtful that valid consent is present, raising the possibility that their RAMI is infringed.

C. Deception

Consider, finally, cases of deception. Here is an example:

> *Deceptive Rain Gesture*. Ada arrives late to a meeting in a windowless room. As she enters the room, she gestures to her soaked clothing, and, on seeing this, and without any mediating conscious thought, Bert comes to believe that it is raining outside, whereas he did not previously believe this. In fact, it's not raining; Ada's clothes are wet because she asked the gardeners to spray her with the hose before entering the building. She did this because she wanted to trick Bert into believing that it's raining.

Is the mediating process that produces Bert's change of belief reasons-insensitive?

One question to consider here is whether the fact that Ada gestured to her wet clothes is itself a reason for Bert to form the belief that it is raining. If it

[21] For more recent reviews of the relevant literature, see Chan and Harris (2017) and Ramnerö et al. (2019).

[22] See, for example, Drummond and Sauer (2018). Randomised rewards are an instance of what is known in the psychological literature as 'variable ratio reinforcement', which has long been known to enhance conditioning processes. See, for example, Ferster and Skinner (1957).

is, it seems clear that the mediating process in this case will come out as significantly reasons-sensitive, for the fact that Ada gestures to her wet clothes clearly plays a substantial causal role in altering Bert's belief, and via the same reasons-tracking kind of process as occurred in the original, non-deceptive *Rain Gesture* case.

Whether the fact that Ada performs the gesture counts as a reason in this case will depend, I think, on whether we allow for reasons to be relativised to an agent's particular epistemic situation. Given full knowledge of the situation, it seems clear that this fact does not constitute a reason to believe that it is raining. But presumably Bert does not have full knowledge of the situation. For example, he presumably does not know the truth about how Ada's clothing got wet. Given what he does know, it seems reasonable for him to form the belief that it is raining. Thus, we might want to say that, *relative to Bert's epistemic situation*, the fact that Ada performs the gesture constitutes a reason for him to believe that it is raining. An alternative view would hold that, given his poor epistemic situation, Bert has an *apparent* reason to believe that it is raining, though no real reason to believe this, since it is, after all, not raining.

I want to remain neutral on whether the fact that Ada performs the gesture counts as a reason for Bert to believe that it is raining, or only an apparent reason. This means I have to remain neutral on whether the mediating process within Bert counts as reasons-sensitive by virtue of its being sensitive to Ada's gesture.

Nevertheless, I doubt that Ada's influence on Bert in this case will satisfy my insensitive process condition. This is because the process mediating Bert's response is plausibly highly sensitive to other facts that *are* reasons bearing on the alteration. Suppose that, in the *Deceptive Rain Gesture* case, Bert's weather app reported a 93% chance of rain in the morning, and that there were dark clouds gathering when he entered the building himself, 15 minutes previously. These facts, which are reasons to form the belief that it is raining, plausibly made a causal contribution to his coming to believe that it is raining. They plausibly played a supporting role, alongside Ada's deceptive gesture. For example, they may have made him more receptive to being deceived by Ada's gesture. This suggests that the first arm of the causal-dispositional condition for reasons-sensitivity is satisfied; reasons significantly causally contributed to the alteration. And the second arm of the condition is plausibly satisfied too: the (implicit) weighing process via which Bert integrates these considerations and comes to believe that it is raining was plausibly a reasons-tracking type of process. It is, after all, the very same

kind of process that was present in the original, non-deceptive *Rain Gesture* case—the two processes share all of their important intrinsic features, and thus, I take it, have the same intrinsic dispositions, including the same intrinsic disposition to produce reasons-conformity. Thus, even if Ada's deceptive gesture itself gives Bert no reason to alter his beliefs, it is still plausible, on my view, that the mediating process was significantly sensitive to other reasons.

I suspect that, in many cases of deception, the mediating process will come out as significantly reasons-sensitive on similar grounds; even if the deceiver creates only an apparent, and not a genuine, reason, the mediating process is likely to be significantly sensitive to a wide range of other reasons. However, it is possible to imagine forms of deception in which this type of reasons-sensitivity is not likely to obtain. Typical cases of influence via subliminal imagery will, I think, be in this category.[23]

Thus, suppose that Odette intentionally causes Parvan to experience a desire to drink Fanta. She does this by setting Parvan's television to display subliminal images of beautiful people drinking Fanta. Though the aptness of this description is not crucial, I find it somewhat plausible to think of this as an instance of deception—Odette causes Parvan to experience the desire by giving Parvan the false impression, albeit a subconscious one, that drinking Fanta makes you beautiful, or at least, more similar, in important respects, to beautiful people. She has presented him with an apparent reason to want to drink Fanta. The mediating process that occurs within Parvan is presumably somewhat sensitive to this apparent reason. However, in contrast to *Deceptive Rain Gesture*, it would be difficult to make the case that this apparent reason is also a genuine reason, even if we allow for reasons that are relativised to a person's epistemic situation; it does not seem reasonable, even given the circumstances, for Parvan to form the belief that drinking Fanta makes you beautiful or more similar, in important respects, to beautiful people. Moreover, Odette has influenced Parvan via a process, in Parvan, that is unlikely to be sensitive to other reasons that bear on the alteration. Parvan could have a range of other pre-existing reasons for or against coming to desire Fanta. For example, the fact that he finds Fanta refreshing might be a reason for him to desire it. The fact that he is a diabetic

[23] The effectiveness of subliminal imagery is a matter of controversy; however, for evidence that it can be affective in altering certain mental states, at least under certain conditions, see, for example, Amd and Passarelli (2020), Karremans et al. (2006), Veltkamp et al. (2011), and Verwijmeren et al. (2011). For an unfortunately now somewhat outdated review of the empirical literature, see Plassmann et al. (2012).

currently experiencing hyperglycaemia may constitute a reason *not* to desire it. But it is doubtful that the mediating process will be sensitive to these reasons. Though it is ultimately an empirical question to which we currently lack an empirically informed answer, I see no theoretical reason to suppose that these facts would exert any influence on the subconscious processes via which subliminal imagery plausibly operates. I think, then, that while the mediating process that occurs within Parvan is likely sensitive to the fact that he has the subconscious impression that drinking Fanta makes you beautiful, or more similar, in important respects, to beautiful people, this is probably not a genuine reason, and it is doubtful that the process is significantly sensitive to other reasons. Thus, I think the process is likely to satisfy the insensitive process condition, making Odette's intervention a potential infringement of his RAMI.

IV. Conclusion

In this chapter, I have explored the implications of my account of the RAMI for a range of different types of mental influence, ranging from aversion therapy to salience-based nudges to nonconsensual neurointerventions. It is, in general, not easy to assess the implications of my account, in part because I have left several aspects of the RAMI unspecified (such as exactly where the threshold for *insufficient* reasons-sensitivity lies, and exactly which considerations count as reasons) and in part because of a lack of relevant empirical evidence: reasons-sensitivity has not attracted much attention from empirical researchers studying the mechanisms via which different types of mental influence operate. Nevertheless, I have sought to identify some types of influence that would be unlikely to infringe the RAMI: reporting reasons, drawing attention to reasons, and creating new reasons. I also identified some types of influence that would be likely to infringe the RAMI, with nonconsensual neurointerventions being the clearest example. And for some 'harder cases', such as salience-based nudges, conditioning-based influences, and deception, I have sought to at least offer some thoughts on why and in which cases these influences would or would not infringe the RAMI. In the course of this discussion, I suggested that conditioning-based influences and the use of subliminal imagery would plausibly infringe the RAMI, at least in some cases.

8

Against the Right Against Mental Interference

Having offered two arguments for the right against mental interference (RAMI), specified its scope, and drawn out some of its implications, in this chapter I turn to consider two objections to the view that we possess this right. Both objections maintain that the RAMI is too broad; it classes too many actions as wrongful. The first objection—considered in Section I—holds that the RAMI is too broad because the mind is extended. The second objection—considered in Section II—holds that, even if the mind is not extended, the RAMI is too broad, since it implies that some everyday and intuitively rather innocuous forms of mental influence are wrongful.

I. The Extended Mind Thesis

One way to argue that the RAMI covers an overly broad range of interventions would be to invoke the extended mind thesis, which I will understand as the view that, at least for some people, certain cognitive aids that are external to the person's body—perhaps including their regularly consulted notebooks, diaries, and smartphones—count among the physical realisers of the person's mental states.[1] Put somewhat less technically, the extended mind thesis holds that the minds of at least some people reside not only in their brains but also, in part, in some of their external cognitive aids.

As many have noted, the extended mind thesis has interesting, and in many cases challenging, ethical implications.[2] Its implications for the RAMI might seem to be among those challenging implications. If the

[1] For the classic defence of this view, see Clark and Chalmers (1998).

[2] For discussion of some of these, see Anderson (2008), Bublitz and Merkel (2014), Buller (2013), Glannon (2014), Levy (2007), and Lippert-Rasmussen (2018).

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0008

extended mind thesis holds, then it looks as though deleting the storage of a person's smartphone or tearing a page out of her frequently consulted diary could count as an infringement of the RAMI, even in the absence of any neural effects; an alteration to the state of those cognitive aids might itself involve mental alterations—the deletion of certain memories, say. Moreover, these mental alterations would seem to have much in common with those in which, say, memories are deleted by electrically suppressing the neural states in which a person's memories are encoded. In both cases, the person's mental states would be altered via a physical intervention on their physical basis. It seems that, if deleting memories by physically altering a person's brain states would infringe the RAMI, then so too would physically wiping the storage of her smartphone or removing a page of her diary. This might be thought to leave the RAMI excessively broad, for though it is intuitively plausible that tampering with a person's smartphone or diary in these ways wrongs them, it is not clearly plausible that it wrongs them *by infringing their RAMI*. It might seem, for example, that the wrong committed in these cases is less grave than it would be if it involved a RAMI-infringement.[3]

I have a quick response and a less quick response to this worry. Here is the quick response: Though, in general, the counterintuitiveness of a moral view counts against it, in this case it may be possible to explain away any counterintuitiveness—that is, to explain it in a way that does not detract from the credibility of the view that we possess the RAMI. It is, after all, hardly surprising that the RAMI will have some counterintuitive implications when combined with the view that our minds are extended, given that the latter view is itself rather counterintuitive. Perhaps, insofar as we find it difficult to see wiping someone's smartphone or removing a page of their frequently consulted diary as an infringement of their RAMI, this is simply because we struggle to see their smartphone or diary as part of the physical basis of their mind. We may be able to appreciate the philosophical arguments for the view that these objects *are* among the physical realisers of a person's mind without being able to really internalise that view in such a way that it guides our moral intuitions regarding how the smartphone or diary should be treated or what

[3] Alternatively, it might be thought implausible that there is a RAMI-infringement here, even leaving aside intuitions about wrongfulness, simply because it is implausible that there is a mental interference. I will not further pursue this form of the worry since I think it is open to the proponent of the RAMI to respond by simply insisting that they are using the term 'mental interference' as a term of art, and are thus not attempting to capture intuitions about what is and what is not mental interference.

right we possess over them.[4] If this is right, then we should not trust our intuitions about such cases.

In reply, it might be argued that this quick response relies on a particular and possibly false story about what is driving our intuitions in the relevant cases. In advancing the quick response, I assumed that our intuitions are being driven by the (if we accept the extended mind thesis, mistaken) assumption that external artefacts, such as our smartphones and diaries, are not really among the physical realisers of our minds. However, there are other possible stories available. One plausible story, for example, is that our intuitions are instead being driven by the thought that the acceptance of an extended RAMI—one that covered such artefacts as our smartphones and diaries—would be problematic because such an extended RAMI would be too demanding; it would be too difficult for moral agents to avoid infringing this right. And that thought, it might be argued, is correct. Moreover, it is correct regardless of whether our smartphones and diaries are among the physical realisers of our minds.

A difficulty with this reply, however, is that, on this alternative story, our intuitions also seem to be grounded on a mistake. For as a matter of fact, even an extended RAMI needn't be particularly demanding. This is because, as I am understanding it, the extended RAMI is not absolute; it corresponds to a *pro tanto* duty, not an all-things-considered one. Thus, it can be permissible to infringe the RAMI, and one type of case in which it would likely be permissible to infringe it is where it would be highly costly to avoid infringing it. This feature of the RAMI powerfully limits its demandingness.

Since I cannot think of another story on which the relevant intuitions are *not* grounded on a mistake, I am inclined to think that my quick response succeeds. But even if it does not, I think it is possible to respond to the worry about counterintuitiveness in a way that rescues the RAMI, or something close to it. This brings me to my less quick response. To explain this response, I need to distinguish two possibilities concerning *how* extended the mind is. The first possibility is that the mind is only *modestly* extended; the external physical bases of our minds include only those few external objects with which our neural processing is especially integrated. (For many of us, our smartphones would perhaps be the strongest candidate for inclusion within this limited set.) The second possibility is that our minds are *radically* extended; the physical bases of our minds

[4] I thank Jonas Hertel and Gabriel De Marco for pressing me to offer this response. For an analogous response to the non-identity problem, see McDermott (1982, p. 166).

include (i) external objects besides those few with which our neural processing is especially integrated, and/or (ii) services (such as Google Search, ChatGPT and Wikipedia), and/or (iii) other people (e.g. those on whom we rely for advice or cognitive support).

I am not at all persuaded that our minds are radically extended, and I do not think many other philosophers are persuaded of this either. But if our minds *are* radically extended, I concede that the RAMI, as I have formulated it thus far, may have very counterintuitive implications. It may, for example, imply that, when Google slightly tweaks their search algorithm, they thereby infringe the RAMI of their billions of users. And for the sake of argument, I will here also concede that the counterintuitiveness of these implications would count against the view that we possess the RAMI. I will assume, in other words, that my quick response above fails. Given these concessions, I think the correct response to the view that our minds are radically extended would be to simply restrict the RAMI so that it covers only *parts* of our minds (perhaps only their non-extended parts, or perhaps their non-extended parts along with *some* of their extended parts). It is important to remember that the external mind thesis is a view about what the boundaries of the mind are, not about where the boundaries of the mind's moral protection should be set. We are free to set the boundaries of that moral protection more narrowly than the boundaries of the mind itself, and if the mind is radically extended, and my quick response above fails, I think we should do just that.

There would, I think, be a legitimate question about whether a right would deserve the name 'right against mental interference' if the right fails to cover some of the mind's externally realised parts. (Recall that, in Chapter 3, I suggested that it is a feature of the RAMI that it covers *all parts of* our minds.) However, I am happy to concede that, if our minds are radically extended, we may *not* enjoy a RAMI, but rather only a right against interference with certain parts of our minds. Anyone who believes that our minds are radically extended should henceforth feel free to read my references to the RAMI as references instead to this more limited right.

In reply to this response, it might be objected that, since the extended parts of our minds are just as much parts of our minds as the non-extended parts, it would be *ad hoc* to exclude (some of) the former from the moral protection offered to the latter. I think it is possible, however, to identify principled grounds for excluding at least some external entities from the scope of the RAMI. For example, though mental states encoded in the brain of a person's carer or friend might perhaps count as a part of the person's radically extended mind, we could perhaps justifiably exclude them from the scope

of their RAMI on the basis that only one person should be assigned a RAMI over some entity, and the friend or carer has a stronger claim to possessing a RAMI over the mental states encoded in their own brain. Similarly, though, say, Google's search engine might perhaps count as part of a person's radically extended mind, we could perhaps justifiably exclude it from the scope of their RAMI on the basis that this would be in tension with Google owning the search engine, and we think there is a stronger basis for ascribing Google a property right over the search engine than for assigning the search engine user a RAMI that covers it.

None of this demonstrates beyond doubt that it is possible to set the boundaries of the RAMI so that it avoids both implausible breadth and *ad hoc* exclusions of some parts of the radically extended mind. There may well be external objects that are (i) plausibly part of a person's (radically) extended mind, but (ii) not plausibly protected by the person's RAMI, although (iii) no principled basis for this exclusion can be given. I can, however, think of no such object. Moreover, I note that, if such objects exist, we may still be able to justifiably endorse a RAMI and exclude these objects from its scope by adopting an *ad hoc* restriction. In ethics, theoretical parsimony and unity do sometimes have to be sacrificed in order to avoid unacceptably counterintuitive implications.

We have been considering the possibility that our minds are radically extended. But what if they are instead extended, but only modestly so? (I am willing to give significant credence to this possibility.) Here too, I *could* respond by allowing that, if my quick response above fails, we should restrict the scope of the RAMI. However, I do not think this is necessary, since I think that combining the RAMI with the view that our minds are modestly extended has implications that are at most mildly counterintuitive, not unacceptably implausible. Thus, I think that if our minds are only modestly extended, we can simply bite the bullet and insist that the RAMI *does* cover the external physical bases of our minds, which may include objects such as our smartphones and diaries.

In fact, I think it is not clear that there is any bullet to bite here. It seems to me that if someone tampers with my smartphone, then she *does* infringe my rights. Moreover, I think there is something to be said in favour of the view that one of the rights she infringes is my RAMI. For it seems to me that the wrong the person has committed against me is more serious than the wrong she would have committed had she interfered with a smartphone that I own, but do not use as a cognitive aid; the fact that I use my smartphone as a cognitive aid plausibly makes a difference to the wrongfulness of the interference

with my smartphone. This could be explained by the fact that it brings my smartphone within the scope of my RAMI.

Let me take stock. In this section, I have considered the worry that, if some version of the extended mind thesis holds, the RAMI will be implausibly broad. I have offered a quick response and a less quick response to this worry. The quick response holds that, since the extended mind thesis is itself out-of-line with our intuitions, the fact that the RAMI would have counter-intuitive implications when combined with that thesis does not count against the existence of the RAMI. My less quick response, which can be deployed if the quick response fails, takes a disjunctive form. If the mind is radically extended, then I suggest that we should constrain the RAMI so that it does not cover the whole mind. On the other hand, if the mind is only modestly extended, I think we can simply insist that the RAMI does cover the extended parts of our minds.

II. Everyday Mental Influences

We have been considering the possibility that the RAMI might be rendered excessively broad by the truth of the extended mind thesis. The extended mind thesis is not, however, the only basis for such worries. Even if our minds are not extended, one might worry that the RAMI will cover a very broad range of actions simply because there are very many rather commonplace ways in which we act on others' minds via processes that are not significantly sensitive to reasons.

Some such cases can perhaps be excluded from the scope of the RAMI on the basis that they do not clearly bring about a deviation from the baseline situation(s), so do not, on my view, clearly involve *altering* someone's mental states. Consider:

> *Warm Colours*. Christoph paints the walls of his restaurant in warm colours—primarily shades of red and yellow—with the goal of thereby enhancing the appetites of his customers. He has read that exposure to warm colours tends to make people feel hungrier. Seeing the walls indeed causes Devi, one of Christoph's customers, to feel hungrier than she would have felt had the walls been a different colour.

In this case, Christoph performs an action that causes Devi's appetite to be stronger than it would have been had the walls been a different colour, and

that, moreover, arguably operates via a reasons-insensitive process within Devi. However, it is not clear that there is any deviation from the baseline situation(s). The state of affairs in which the walls are painted in warm colours might well have come about for any number of other reasons and could certainly be aptly characterised as a 'normal' or 'expected' situation. I thus think it has a good claim to be among the range of baseline situations. Thus, this case does not clearly involve *altering* mental states, as I am using that term, and so does not clearly constitute a mental influence, as I am using that term, let alone a mental interference.

There are, however, other seemingly innocuous actions that it is more difficult to exclude from the scope of the RAMI. Consider:

> *Perfume.* Before going on a date with Beau, Aroha applies a perfume that is known (including by Aroha) to be unusually powerful in its ability to produce feelings of attraction in those who smell it. Beau smells the perfume, and this indeed causes him to experience feelings of attraction for Aroha.

This is a clearer example of a mental influence. Aroha performs an action that alters Beau's feelings of attraction compared to what is plausibly the baseline situation, in which Aroha does not wear perfume, or wears a perfume that is not unusually powerful.[5]

So I suppose that Aroha's action of wearing perfume alters Beau's feelings of attraction, and thus that she mentally influences Beau. Does she also mentally *interfere* with him? On my account, she does if the mediating process within Beau is insensitive to the reasons that bear on the alteration that it produces. It seems possible that this process is indeed reasons-insensitive. Certainly, we can refine the case so as to make this plausible. Thus, suppose that the perfume that Aroha wears in this case is one that often strongly enhances feelings of attraction in those who smell it. Suppose moreover that the perfume does not confer on the perfume-wearer any property that might plausibly be thought to make those feelings apt—such as the property of smelling more pleasant. Nor, in this case, does Aroha's wearing of the

[5] In fact, though I think it is plausible, I do not think it is obvious that there is a deviation from the baseline situation here. Perhaps wearing an unusually powerful perfume counts as 'normal', and thus as among the set of baseline situations. Nevertheless, for the sake of argument, I will grant that wearing the unusually powerful perfume on a date constitutes a deviation from the baseline. If you find this hard to believe, it may help to suppose that Aroha and Beau live in a society in which wearing any perfume at all is rare.

perfume indicate or draw Beau's attention to the presence of any plausibly desirable feature of Aroha—such as the possession of good aesthetic taste. Nor, moreover, do independent reasons make any causal contribution to the attraction-boost produced by the perfume. For example, it is not the case that the perfume's attraction-boosting effect is enhanced by Aroha's possession of various unrelated properties that Beau desires in a romantic partner, nor that it is diminished or restrained by her lacking other properties that Beau desires. Nevertheless, the perfume does work; it does enhance Beau's feelings of attraction for Aroha.

Given these further assumptions, it is hard to see how the mediating process that occurs within Beau could be significantly sensitive to the reasons that bear on the attraction-boost that the perfume in fact causes. Thus, it seems that Aroha's intervention will come out as a mental interference, on my view. And this means that Aroha may have infringed Beau's RAMI. (She will have infringed it if it is also the case that (i) the interference counts as significant, (ii) Beau possesses the RAMI, and has not waived or forfeited it in relation to Aroha's action, e.g. through consenting to it.) But it might be thought implausible that Beau's RAMI is infringed in this case.[6]

How can a defender of the RAMI respond to the *Perfume* case?

One option would be to maintain that the interference in this case is too minor to qualify as an infringement of the RAMI; it does not satisfy the significance condition. However, since feelings of romantic attraction plausibly count as rather important mental states, I do not find this response plausible, at least if we suppose—as I henceforth will—that the new feelings of attraction produced by the perfume are rather strong.

Alternatively, one could maintain that the interference in *Perfume* does not infringe the RAMI because it is consensual. Though it is doubtful that Beau will have explicitly consented to Aroha's intervention, it might be

[6] Gibert (2023, pp. 339–47) discusses a wider range of possible counterexamples to my claim that we possess a RAMI. She discusses a range of cases that, she suggests, constitute 'non-rational' influences on practical reasoning but are not, intuitively, wrongful. However, many of the influences that Gibert invokes would not, on my view, clearly constitute mental interferences, since they operate by implicitly drawing a person's attention to reasons (examples include 'exposing [people] to new experiences' (p. 340), 'sending one's teenage children to a diverse public high school to instill in them respect for those from different backgrounds' (p. 341), and 'showing a friend pictures of lung disease to help them stop smoking' (p. 341)) or, like my *Warm Colours* case, do not clearly involve a deviation from the baseline ('painting the dentist's office pink to calm patients' (p. 341)). Her other examples are, I think, rather similar to my *Perfume* case, so fairly represented by it. Indeed, Gibert explicitly discusses the case of 'wearing perfume on a date' (p. 341).

argued that, in agreeing to go on a date, one implicitly consents to being exposed to attraction-enhancing perfumes and therefore waives one's RAMI in relation to such exposures.

This response may succeed on some further specifications of *Perfume*, but I doubt that it succeeds in relation to *all* variants of this case. Valid consent is normally taken to require that the consenter is, or at least should be, aware of the relevant features of the intervention to which consent is given. Yet we can suppose that Beau was not aware that perfume can enhance feelings of attraction. Moreover, we can suppose that this ignorance was perfectly reasonable; perhaps there is not much information regarding these effects in the public domain and perhaps it is relatively difficult to acquire this information. These further suppositions may cause some readers to get off the boat, concluding that, if we accept them, *Perfume* does start to look like a case of wrongful mental interference. However, I suspect some will continue to intuit that Aroha has not wronged Beau. Yet it is now clear that Beau has not validly consented to the influence.[7] Thus, my account implies that Beau *has* been wronged in this variant of the case.[8]

What further responses are available to the defender of the RAMI? I can see two. The first would be to refine the RAMI in such a way as to exclude *Perfume* from its scope. The most promising way to do this would, I think, be to limit the RAMI so that it does not cover cases in which the action of the influencer is protected by some special prerogative. It could then be argued that Aroha's action in *Perfume* is protected by such a prerogative.

[7] It might be argued that, even if Beau has not validly consented to Aroha's wearing of perfume, since he is not sufficiently informed about its effects, he has validly consented to a broader class of actions of which Aroha's wearing a perfume is a member. Everyone knows that when you go on a date, you might be exposed to influences—like fancy clothes, carefully applied make-up, and charm—that may boost one's feelings of attraction, including via processes in the influencee that fall short, in various ways, of the ideal of rational thought. Perhaps when Beau agrees to the date, he implicitly consents to being subjected to any and all actions that belong to this class of potentially attraction-boosting influences. And perhaps this belief is sufficiently well-informed to be valid; even if Beau is mistaken about the attraction-boosting effects of perfume, his beliefs about the broader class of potentially attraction-boosting influences might be close enough to being correct. Thus, perhaps Beau has validly consented to Aroha's action *qua* potentially attraction-boosting influence, but not *qua* wearing of attraction-boosting perfume. To exclude this possibility, we would have to modify the case such that his implicit consent to the broader class of actions is also invalid, say, because it was insufficiently well-informed. Perhaps, across the board, Beau dramatically—and reasonably, given his epistemic situation—underestimates the attraction-boosting effects of potentially attraction-boosting influences. Even in this variant of the case, it seems to me doubtful that Beau has been wronged, thus this variant of the case would remain a potential counterexample to the RAMI, as I have formulated it. I thank Sasha Arridge for pressing me to consider this issue.

[8] For an argument similar to the one I make in this paragraph, see Gibert (2023, p. 342).

For example, it could be claimed that Aroha enjoys a prerogative to 'wear' what she likes, and that this takes her action beyond the scope of Beau's RAMI.

Yet again, however, I do not find this response compelling. Even if Aroha enjoys this prerogative, or something like it, I am not convinced that this will prevent her from infringing Beau's RAMI.

In general, when prerogatives and rights bump up against one another, there are three possible ways to reconcile them: allow the prerogative to carve out an exception in the right, allow the right to carve out an exception in the prerogative, or accept that the right and the prerogative are in conflict and must be balanced against one another. To illustrate, suppose that we each enjoy a free speech prerogative but also a right against being the object of hate speech. In cases where a speaker subjects a listener to hate speech, this prerogative and right bump up against one another, and there are three obvious ways in which we could reconcile them. We could say that the listener's right against hate speech does not cover cases where the speaker enjoys a free speech prerogative, thus allowing the speaker's prerogative to carve out an exception in the listener's right. We could say that the speaker's free speech prerogative does not cover cases where the speech is hateful, thus allowing the right to carve out an exception in the prerogative. Or we could simply accept that there is a conflict between the prerogative and the right and that the two must be balanced.

In cases such as these, it is not always clear that the prerogative should carve out an exception in the right. Indeed, in the hate speech case, I and many others find it more plausible that the right carves out an exception in the prerogative.[9] Thus, even if Aroha enjoys a prerogative to wear what she likes, it may not follow that we should restrict the scope of the RAMI to accommodate this prerogative.

We were interested in whether, by appealing to a special prerogative, we might be able to restrict the scope of Beau's RAMI such that it is not infringed in *Perfume*. Thus, the important question for us is whether Aroha enjoys what I will call an *undercutting* prerogative, where by that I mean a prerogative that does carve out an exception in the RAMI.

To explore the possibility that she does, it will be helpful to distinguish two quite different ways in which we could understand Aroha's prerogative. When we say that Aroha enjoys a prerogative to wear what she likes, this could mean either of two quite different things. (I am assuming here that it

[9] See, for example, MacKinnon (1993), Parekh (2012), and Waldron (2014).

is a moral, rather than a legal prerogative.) It might mean that she enjoys a moral right against interference with what she wears.[10] Alternatively, it might be that she enjoys a moral permission to wear what she likes—that is, that there is no *pro tanto* duty for Aroha not to wear any particular item.[11] I think it is very doubtful that Aroha enjoys a prerogative of the latter kind. Consider a chemical, *C*, that, at room temperature, evaporates into a mildly poisonous gas. Suppose that you have some of this chemical, currently stored in a tightly sealed bottle. It seems that you have a *pro tanto* duty not to expose others to this chemical. Thus, for example, it seems that you have a *pro tanto* duty not to open the bottle in the vicinity of others. And it seems that this *pro tanto* duty would remain in place if the mode of exposing others to the chemical was to 'wear' the chemical on one's own skin. That this is the mode of exposure does nothing to weaken the *pro tanto* duty. And I think this point generalises. Our *pro tanto* duties to others are not eliminated or weakened in cases where the putatively duty-violating action involves wearing something.

It is more plausible that Aroha possesses a prerogative of the former kind, that is, a right against interference with what she wears (though this is clearly not an absolute right). But note that this sort of prerogative would do nothing to prevent Aroha from infringing Beau's RAMI by wearing the perfume. It does not follow from the fact that one enjoys a right against interference with one's ϕ-ing that one would infringe no-one else's rights by ϕ-ing.[12] Possessing a prerogative of this kind is quite consistent with it being the case that, in exercising one's prerogative, one infringes the right of another in the sense of 'right' that I have been employing throughout this book: one fails to live up to some *pro tanto* duty owed to the other. It is thus unclear why a prerogative of this kind would carve out an exception in the RAMI.

Given these difficulties with the appeal to special prerogatives, I am inclined to prefer a different response to the problem posed by *Perfume*. My preferred response is to simply bite the bullet and concede that Aroha does wrong Beau in *Perfume*, given the further stipulations that Beau is unaware, and reasonably so, of the attraction-enhancing effects of perfume, and that the feelings of attraction produced by the perfume are strong.

[10] In Hohfeldian terms, this would be a claim-right.

[11] In Hohfeldian terms, this would be a privilege.

[12] Here I am assuming, as I have been throughout, that moral rights need not be enforceable. Thus, it does not follow from the fact that some action of mine would infringe another's rights that the other person can, without infringing my rights, interfere with my performance of this action. I could have a right that would be infringed by this (attempt at) enforcement even though my action did infringe her right, in the sense that it failed to fulfil some *pro tanto* duty that I owed to her.

We can, I think, render this bullet more palatable than it might otherwise seem by allowing that mental interference is a graded wrong; it comes in various degrees of wrongfulness, depending (perhaps among other factors) on the magnitude of the interference. Even when a mental interference is large enough in magnitude to meet the significance condition, and thus to infringe the RAMI, the seriousness of the infringement might depend on how far above the threshold for significance the interference lies. When the magnitude of the interference is great enough to count as significant, but not *that* great, the infringement of the right may not be *that* serious.

This would plausibly mirror the situation with bodily interference.[13] Most of us would think that, though giving someone a gentle shove constitutes bodily interference, and bodily interference of sufficient magnitude to infringe the right against bodily interference, it is a lesser interference than, for example, cutting a person's skin with a knife, which in turn is a lesser interference than, say, amputating a person's limb. These differences in the magnitude of the bodily interference plausibly give rise to a difference in wrongfulness. Other things being equal, it is more wrongful to amputate a person's limb without their consent than to cut their skin without their consent, which is more wrongful than to give them a gentle shove without their consent.

If we allow that, like infringements of the right against bodily interference, infringements of the RAMI come in different degrees of seriousness, we are not committed to holding that wrongful mental interference is always seriously wrongful. Moreover, in relation to the *Perfume* case specifically, it may be plausible to hold that Aroha's action is not seriously wrongful because the mental interference that she performs is not large in magnitude. After all, she may interfere with only one mental state—Beau's feelings of attraction for her—on only one occasion. To be clear, the suggestion here is that the mental interference may be large enough in magnitude to count as 'significant', and thus, on my view, to infringe the RAMI, but not large enough to count as a (very) serious infringement of the RAMI.

I personally find it plausible—or at least, not unacceptably implausible—that Aroha *does* mildly wrong Beau in my further specified version of *Perfume*. (Recall my stipulations that Beau is unaware, and reasonably so,

[13] See, for statements of the view that infringement of the right against bodily interference comes in degrees of seriousness, Archard (2008, p. 28), Weinrib (1995, p. 128), and Wertheimer (2003, p. 109). For defence of the view that the *legal* right against bodily interference is also graded, see Herring and Wall (2017, p. 573).

of the attraction-enhancing effects of perfume, and the feelings of attraction produced by the perfume are strong.) However, I accept that some might find it counterintuitive that there is any wrong at all in this case. They may find it more plausible that Aroha's intervention is entirely innocuous, morally speaking. Thus, I accept that endorsing the RAMI, as I have specified it, may come at some cost to our ability to accommodate the intuitions of some. However, in earlier chapters, I have argued that endorsing the RAMI also comes with some *benefits* to our ability to accommodate intuitions. Recall, for example, the *Dream Modulation* case from Chapter 4, in which you render Lucien's drug-induced dreams less gloomy by administering a further drug. I argued that endorsing the RAMI allows us to capture the intuition that you wrong Lucien in the *Dream Modulation* case—an intuition that is, I suggested, difficult to accommodate in other ways. Since I think it is harder to accept that there is no wrong in *Dream Modulation* than it is to accept that there is a mild wrong in the further specified version of *Perfume*, I think we would do better to accept the RAMI than to give up on explaining the wrong in *Dream Modulation*.

III. Conclusion

In this chapter, I have considered two objections to my claim that we possess a RAMI. Both of these maintained that the RAMI is implausibly broad: it implies that some intuitively innocuous interventions are in fact morally wrongful. The first objection appealed to the extended mind thesis and maintained that the RAMI will, if the extended mind thesis holds, be infringed by interferences with the external objects in which our minds are partly realised. I first offered a quick response to this objection, according to which our intuitions regarding interferences with external cognitive aids can be dismissed on the basis that they reflect our failure to fully internalise the extended mind thesis. I then offered a less quick response that can be employed if the quick response fails. I conceded that, if our minds are *radically* extended, we will need to restrict the scope of the RAMI so that it no longer covers our *whole* minds. But I also argued that, if our minds are only *modestly* extended, we can plausibly maintain that the RAMI does cover our whole minds, and, thus, that some external cognitive aids fall within its scope.

The second objection maintained that our possession of the RAMI would implausibly imply that some rather mundane and intuitively unproblematic interventions, like that in *Perfume*, would infringe the right. I conceded that

the intervention present in a further specified version of *Perfume* would indeed likely infringe the RAMI, as I have understood it. However, I sought to render this implication palatable by arguing that the RAMI is a graded right—one that can be more or less seriously infringed—and by citing grounds for thinking that any infringement present in the refined version of *Perfume* would lie at the less serious end of the spectrum.

There are no doubt other objections to the RAMI that I have not considered, but the two objections considered in this chapter are, it seems to me, the most troubling ones. Thus, I hope that in responding to them, I have increased the credibility of my claim that the RAMI is a right that we possess.

9
Concluding Thoughts

In this book, I have outlined and defended a moral right against mental interference (RAMI), taking as my starting point the more widely endorsed moral right against bodily interference (RABI). In this final chapter, I summarise my arguments (Section I) and offer some comments on their limitations and possible further implications (Section II).

I. Summary of the Arguments

In Chapter 1, I introduced and motivated the thought that we possess the RAMI, where that right was understood as a mental analogue of the better understood and more widely accepted RABI. I also explained why I think the RAMI warrants greater attention than it has so far received: the existence of the right has considerable *prima facie* plausibility, the mind is highly vulnerable to interference (suggesting that the RAMI might be frequently infringed), and a study of the moral RAMI would help us to assess the advisability of recent moves to recognise similar *legal* rights.

In Chapters 2 and 3, I offered schematic, tentative, incomplete, and metaphysically ecumenical working accounts of the rights against bodily and mental interference. These accounts specified conditions that I took to be jointly sufficient for each of these rights to be infringed.

On my working account of the RABI, it is sufficient for *A* to infringe *B*'s RABI that *A* is a well-situated moral agent, *A* significantly interferes with *B*'s body, and *B* possesses the RABI and has not waived or forfeited it in relation to the interference in question.[1] Further, it is sufficient for *A* to interfere with

[1] Recall that the 'well-situated' implies that there is nothing to prevent the agent from bearing the duty that corresponds to the RAMI in the particular case in question. I take it that a moral agent will count as 'well-situated' if (i) they could reasonably be expected to foresee the consequences or other features of their conduct by virtue of which they interfere with the putative right-holder's mind in a particular case, and (ii) there is nothing severely impeding their exercise of their cognitive and moral capacities.

Protecting Minds. Thomas Douglas, Oxford University Press. © Thomas Douglas 2026.
DOI: 10.1093/9780191979651.003.0009

B's body that *A* (i) touches *B*'s body or (ii) performs an action that alters at least one of *B*'s bodily states via a mediating process, in *B*, that is insensitive to the reasons that bear on the alteration.

Similarly, on my working account of the RAMI, it is sufficient for *A* to infringe *B*'s RAMI, that *A* is a well-situated moral agent, *A* significantly interferes with *B*'s mind, and *B* possesses the RAMI and has not waived or forfeited it in relation to the interference in question. And it is sufficient for *A* to interfere with *B*'s mind that *A* performs an action that alters at least one of *B*'s mental states via a mediating process, in *B*, that is insensitive to the reasons that bear on the alteration.

In Chapters 4 and 5, I made a case in favour of the existence of the RAMI, thus conceived. Chapter 4 developed an intuitive, abductive argument for the RAMI. There, I introduced the *Brain Drug* case, in which you attenuate Lucien's gloominess by spraying a brain-active drug onto his hand. I argued that we cannot fully and satisfyingly account for the wrongfulness of your actions in this case and a range of variants of it by appealing to the claims that you treat Lucien in a way that he has not autonomously chosen, interfere with his body, reduce his mental autonomy, or unduly influence him—at least, not unless we understand those claims in a way that implies a RAMI. Chapter 5 offered a theoretical defence of the RAMI—the argument from self-ownership—which derived the claim that we possess a RAMI from the claim that we own ourselves. I motivated this defence, and defended it against three objections, arguing that none of these objections is fatal to it.

In Chapter 6, I turned from questions of justification to questions of scope. I sought to develop the working account of the RAMI developed in Chapter 3 by further specifying two conditions contained in this account: the significance condition and the insensitive process condition.

According to the significance condition, a well-situated moral agent's interference with the mind of a nonliable individual infringes the RAMI if the interference is significant, in the sense that the magnitude of the interference crosses some threshold. I did not attempt to specify the location of this threshold, but I did identify a range of factors that plausibly affect the magnitude of a mental interference and thus whether the threshold is crossed: the importance of the mental states affected, the number of mental states affected, and the magnitude of the alterations made to them.

According to the insensitive process condition, a mental influence constitutes a mental interference if the mediating process—the process within the influencee via which the alteration is produced—is insensitive to the reasons that bear on the alteration. While seeking to remain neutral on the

main theoretical disputes concerning reasons-sensitivity, I offered a sufficient condition for a mediating process to qualify as reasons-insensitive: the causal-dispositional condition. On this condition, a mediating process is reasons-insensitive if either (i) reasons did not significantly causally contribute, via that process, to the resulting alteration, or (ii) the process was not of a reasons-tracking type—that is, a type of process that is significantly and intrinsically disposed to translate reasons of the sort that were causally operative into mental alterations that conform to those reasons.

Chapter 7 built on Chapter 6 by exploring the implications of my account of the RAMI. I sought to identify some types of mental influence that would be unlikely to infringe the RAMI (influences that consist in reporting reasons, implicitly drawing attention to reasons, or creating new reasons) and others that would be likely to do so (most notably, nonconsensual neurointerventions). I also discussed some hard cases—cases in which either the implications of my account are unclear, or it is unclear whether the implications are plausible. In relation to these cases, I sought to at least offer some preliminary thoughts on whether and why these influences would or would not infringe the RAMI. Two types of influences that, I suggested, could plausibly infringe the RAMI were conditioning-based influences and influences employing subliminal imagery.

Finally, in Chapter 8, I considered two objections to the claim that we possess the RAMI, both of which sought to establish that the RAMI, as I have formulated it, is implausibly broad: it implies that some mental influences infringe the right when intuitively they do not. In relation to some cases—most notably some variants of my *Perfume* case—I conceded that the RAMI may have implications that some would regard as counterintuitive. However, I sought to show that, once we allow that RAMI-infringements vary in their degree of seriousness, we can maintain that these implications are not *that* counterintuitive. I also noted that positing the RAMI allows us to avoid counterintuitive implications in other cases, such as my *Dream Modulation* case from Chapter 4.

II. Limitations and Further Implications

I hope that my arguments will contribute to putting the RAMI on the moral map, and to stimulating further work on this topic. I also hope that at least some readers will be convinced that we do possess this right. My arguments do have some very serious limitations, however.

First, my account of the RAMI is in several respects incomplete. One such respect is obvious: my account offers only a set of jointly sufficient conditions for infringing the RAMI, not a set of necessary conditions. I think my account captures one important class of RAMI-infringing mental interferences, but it leaves open the possibility that there are other ways of infringing the right that are not captured by my conditions. For example, there may be ways of infringing the RAMI that operate via highly reasons-sensitive processes in the influencee, and so do not come out as RAMI-infringements on my account. Coercive mental influences are a possible example. When one person coerces another into desiring this or intending that, she typically creates and makes obvious a very strong reason for that other person to form the desire or intention, and the mediating process may be highly sensitive to that reason. Yet, even when the mediating process is so-sensitive, it is not obviously mistaken to think of coercive mental influences as form of wrongful mental interference.

My account is also incomplete in other ways. For example, I have left unanswered various questions concerning what it takes to alter a person's mental states, such as how to determine what counts as the 'baseline' for the purpose of identifying alterations to a person's mental states. I also did not offer a full account of the magnitude of an interference, of what magnitude a mental interference must have to count as 'significant', and of what, besides the magnitude of an interference, might determine the seriousness of an infringement of the RAMI.

Another limitation of my discussion is that I have said nothing about *who* possesses the RAMI—I have referred to the holders of this right using imprecise and undefined terms like 'we', 'us', and 'persons'. I take it that all adult human beings who are not severely cognitively impaired would possess the RAMI, if anyone does, but beyond that, things become unclear. Do children possess the right? If so, from what age? Do severely cognitively impaired adult human beings possess it? What about non-human animals? Could artificial intelligences possess it? All interesting questions, none of which I have answered.

The important role played by reasons-sensitivity in my account of the RAMI does suggest one possible limit on who could possess the right. It might seem that a being would have to possess the capacity for reasons-sensitivity in order for it to make sense to require—as the RAMI in effect does—that mental influences operate via reasons-sensitive processes in the influencee.

However, even then, there is scope for disagreement regarding whether *all* beings that possess the capacity for any degree of reasons-sensitivity possess the RAMI, or only some subset thereof. As I have been understanding reasons-sensitivity, it comes in degrees, and it may come very cheaply. It is certainly compatible with everything that I have written in this book that very young children and many non-human animals exhibit a significant degree of reasons-sensitivity. Do they thus possess a RAMI? Perhaps. But this could be disputed. For example, it may be that only beings with some minimal capacity to exercise their rights—for example, through demanding that they be respected, or waiving them through consent—could possess the RAMI. Or it may be that some form of conception of oneself is required for possessing the RAMI, perhaps on the basis that self-ownership inherits its value in part from this conception. I leave these as matters to be investigated in further work.

Investigating when children acquire the RAMI is, I think, an especially interesting and important task, since quite a few common parenting techniques arguably constitute mental interferences on my account: consider, for example, the use of dim lighting or a warm cup of milk to encourage sleepiness in one's child. Intuitively, these practices seem unproblematic, but if the children subjected to them already possess the RAMI, we might have to concede either that they are at least mildly *pro tanto* wrong, or that parents must have the power to waive a child's RAMI on her behalf.

A final important limitation of my discussion concerns my strategy of taking the right against *bodily* interference as my point of departure. At some points, I have taken the RABI as a fixed starting point; I have offered arguments for the RAMI that rely on the assumption that we possess a RABI. But in fact, the existence of the RABI can be, and has been, questioned.[2] If we do not in fact possess a RABI, then some strands in my argument for the RAMI will unravel (although others, such as my intuitive, abductive argument from Chapter 4, will not).

In fact, some readings of my own arguments might themselves call the RABI into question. Suppose one is convinced that, if we possess a RABI, then we must also possess a RAMI, perhaps on the basis that the RABI most plausibly derives from something like self-ownership, which would also imply a RAMI. But suppose that one finds the counterintuitive implications

[2] See, for example, Lippert-Rasmussen (2008, pp. 86–118), though note that Lippert-Rasmussen is here rejecting the existence only of an especially strong and robust type of right against bodily interference.

of the RAMI in some variants of *Perfume* unacceptable. Then one might reason as follows: the RABI commits us to the RAMI, which commits us to unacceptable verdicts on some cases; we should therefore reject the RABI. I think this is an interesting line of argument, and I have no objection to its logic; however, I don't myself find the implications of the RAMI so hard to accept.

Further work on the RAMI could seek to address all of the limitations of my arguments discussed above. It could also extend the discussion in other ways.

For example, it could explore the relationship between mental interference and some other, perhaps related or proximate concepts, such as manipulation, brainwashing, and mind control. I have touched on these relationships at various points. For example, I have suggested that wrongful mental interferences need not be controlling nor—at least on some influential accounts of manipulation—manipulative. But I have not comprehensively analysed these relationships.

Further work could also, of course, seek to draw out the possible legal implications of the discussion. As noted in Chapter 1, there have been recent calls for the recognition of new legal rights against certain forms of mental interference, or at least for fuller specification of the ways in which existing legal rights already protect against mental interference. There have also been legislative moves in this direction. These have been taken to their greatest extent in Chile, where, in 2021, the Senate approved a constitutional amendment creating a set of so-called neurorights[3] intended to protect against objectionable uses of neurotechnology to modify or access brains and minds (Cornejo-Plaza & Saracini 2023; Guzmán 2022; McCay 2022), among which is a requirement to 'respect . . . mental integrity' (Fins 2022, p. 3), and in Spain, where, in the same year, the government adopted a 'Charter of Digital Rights' which is intended to inform the development of future regulations and legislation, and which specifies 'psychological integrity' as a permissible basis for regulating neurotechnologies (Government of Spain 2021§ 5.XXVI, 1(d)). A number of other countries, including Argentina, France, Mexico, and the USA are also considering similar moves (Do et al. 2024; Guzmán 2022; López-Silva, 2025).

Scholarly discussion of these (prospective) legislative moves has largely focussed on the question whether they are necessary—some argue that

[3] The term was introduced by Ienca and Andorno (2017).

existing legal rights already offer adequate protection[4]—and whether they might have unintended negative consequences.[5] But a further question is whether these moves should have been tied as closely as they have been to neurotechnologies specifically. The analysis in this book suggests that neurotechnological interventions are not unique in their ability to infringe our moral RAMI. Other interventions—perhaps including hypnosis, subliminal imagery, and various forms of conditioning—could also infringe that right.[6] Thus, any legal reforms intended to enforce this moral right would need to be in one way broader than existing proposals for neurorights, which have focussed specifically on regulating neurotechnologies.

I should acknowledge, however, that the legal implications of the RAMI are far from clear. There is no straightforward route from this moral right to an analogous legal right. To mention only one of the many issues here: there are constraints on the recognition of legal rights that do not apply to moral rights. Legal rights must arguably be such that they can be effectively enforced, and at reasonable cost, and such that the sanctions that might accompany their infringement are predictable.[7] It is not clear that either constraint applies, or applies so strongly, to moral rights (which is why I have not much concerned myself with questions of enforceability or predictability in this book), but both constraints would bite hard when it comes to considering a legal right against mental interference; problems concerning enforceability and predictability in relation to such a right would, I suspect, be severe.

It thus remains to be determined whether a moral right against mental interference would have any interesting legal implications and if so, what they might be. However, I do maintain that the bare existence of such a moral right would in itself, and leaving aside the law, be of practical significance. What practical significance would it have? This is itself a topic for further work, but I will end by making a tentative suggestion: I think my account of the RAMI may suggest that many rather prevalent forms of mental influence are also potentially more problematic, morally, than has often been thought. Examples here might include conditioning-based techniques of

[4] See, for example, Bublitz (2022), Hertz (2023), and Ligthart, Bublitz et al. (2023).

[5] See, for example, Fins (2022).

[6] Jan Christoph Bublitz (2022) makes a more general version of this point.

[7] On the issue of predictability, the European Court of Human Rights understands the principle of legal certainty to require that 'the law should be accessible to the persons concerned and formulated with sufficient precision to enable them—if need be with appropriate advice—to foresee, to a degree that is reasonable in the circumstances, the consequences which a given action may entail' (*Maestri v Italy* 2004, para 30). For a survey of other definitions of legal certainty, see Fenwick et al. (2017, pp. 2–9).

influence, like the use of randomised rewards in computer games, and some forms of behavioural therapy, such as systematic desensitisation. To date, interventions of this kind have been seen as potentially somewhat problematic, but typically less so than nonconsensual medical procedures, even medical procedures that are low risk and minimally invasive. This can be seen, for example, in social expectations regarding consent. Whereas it is normally expected that well-informed and typically explicit consent should be obtained for even safe and minimally invasive medical procedures, like flu vaccinations and mole-excisions, the expectations concerning consent in relation to, for example, loot boxes in computer games or even conditioning-based forms of behavioural therapy appear to be much weaker. When consent is obtained for such interventions, it is very often neither explicit nor informed.[8] This may suggest that many regard the nonconsensual deployment of such mental influences as no more than slightly problematic. I have not directly contradicted that view in this book—since I have not sought to quantify the wrongfulness of different mental interferences—however, my arguments for the existence of a RAMI raise the possibility that they are more problematic than is commonly assumed, and provide us with some new resources to explain why.

[8] For a defence of the view that prevailing consent practices in relation to psychological therapies are less stringent than those regarding bodily interventions, see Forsberg et al. (2025). See also Beahrs and Gutheil (2001) and Trachsel et al. (2015).

References

Alegre, S. (2017). 'Rethinking Freedom of Thought for the 21st Century', *European Human Rights Law Review*, 3: 221–33.

Alex, Wiltshire. (2017). 'Behind the Addictive Psychology and Seductive Art of Loot Boxes', *PC Gamer*. Retrieved 30 November 2025 from https://www.pcgamer.com/behind-the-addictive-psychology-and-seductive-art-of-loot-boxes/.

Amd, M. & Passarelli, D. A. (2020). 'Dissociating Preferences from Evaluations Following Subliminal Conditioning', *Acta Psychologica*, 204: 103023. DOI: 10.1016/j.actpsy.2020.103023

Anderson, J. (2008). 'Neuro-Prosthetics, the Extended Mind, and Respect for Persons with Disability'. Düwell, M., Rehmann-Sutter, C., & Mieth, D. (eds) *The Contingent Nature of Life: Bioethics and Limits of Human Existence*, International Library of Ethics, Law, and the New Medicine, Vol. 39, pp. 259–74. Springer. DOI: 10.1007/978-1-4020-6764-8

Archard, D. (2008). 'Informed Consent: Autonomy and Self-Ownership', *Journal of Applied Philosophy*, 25(1): 19–34. DOI: 10.1111/j.1468-5930.2008.00394.x

Arneson, R. J. (2011). 'Side Constraints, Lockean Individual Rights, and the Moral Basis of Libertarianism'. Bader, R. & Meadowcroft, J. (eds) *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, pp. 15–37. Cambridge University Press.

Arpaly, N. (2003). *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.

Austin, J. L. (1975). *How to Do Things with Words* (J. O. Urmson & M. Sbisà, Eds), 2nd ed. Harvard University Press.

Barton, A. & Grüne-Yanoff, T. (2015). 'From Libertarian Paternalism to Nudging—and Beyond', *Review of Philosophy and Psychology*, 6(3): 341–59. DOI: 10.1007/s13164-015-0268-x

Beahrs, J. O. & Gutheil, T. G. (2001). 'Informed Consent in Psychotherapy', *American Journal of Psychiatry*, 158(1): 4–10. DOI: 10.1176/appi.ajp.158.1.4

Bennett, J. (1993). 'Negation and Abstention: Two Theories of Allowing', *Ethics* , 104(1): 75–96.

Bhargava, V. R. & Velasquez, M. (2021). 'Ethics of the Attention Economy: The Problem of Social Media Addiction', *Business Ethics Quarterly*, 31(3): 321–59. DOI: 10.1017/beq.2020.32

Blanco-Duque, C., Chan, D., Kahn, M. C., Murdock, M. H., & Tsai, L. (2024). 'Audiovisual Gamma Stimulation for the Treatment of Neurodegeneration', *Journal of Internal Medicine* , 295(2): 146–70. DOI: 10.1111/joim.13755

Blitz, M. J. (2010). 'Freedom of Thought for the Extended Mind: Cognitive Enhancement and the Constitution', *Wisconsin Law Review*, 2010(4): 1049–117.

Blitz, M. J. & Bublitz, J. C. (Eds). (2021). *The Law and Ethics of Freedom of Thought. Volume 1: Neuroscience, Autonomy, and Individual Rights*. Palgrave Studies in Law, Neuroscience, and Human Behavior. Palgrave Macmillan.

Blumenthal-Barby, J. S. (2012). 'Between Reason and Coercion: Ethically Permissible Influence in Health Care and Health Policy Contexts', *Kennedy Institute of Ethics Journal*, 22(4): 345–66.

Blumenthal-Barby, J. S. (2014). 'A Framework for Assessing the Moral Status of "Manipulation"'. Coons, C. & Weber, M. (eds) *Manipulation: Theory and Practice*, pp. 121–34. Oxford University Press. DOI: 10.1093/acprof:oso/9780199338207.003.0006

Blumenthal-Barby, Jennifer S. (2021). *Good Ethics and Bad Choices: The Relevance of Behavioral Economics for Medical Ethics*. MIT Press.

Blumenthal-Barby, J. S. & Naik, A. D. (2015). 'In Defense of Nudge–Autonomy Compatibility', *American Journal of Bioethics*, 15(10): 45–47. Taylor & Francis. DOI: 10.1080/15265161.2015.1074304

Bratman, M. E. (2005). 'Planning Agency, Autonomous Agency'. Taylor J. S. (ed.) *Personal Autonomy*, 1st ed., pp. 33–57. Cambridge University Press. DOI: 10.1017/CBO9780511614194.002

Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Luistro Jonsson, M., & Rahali, B. (2018). 'Can Nudges Be Transparent and Yet Effective?', *Journal of Economic Psychology*, 65: 41–59. DOI: 10.1016/j.joep.2018.02.002

Bublitz, J. C. (2014). 'Freedom of Thought in the Age of Neuroscience: A Plea and a Proposal for the Renaissance of a Forgotten Fundamental Right', *Archiv für Rechts- und Sozialphilosophie*, 100(1): 1–25. DOI: 10.25162/arsp-2014-0001

Bublitz, J. C. (2022). 'Novel Neurorights: From Nonsense to Substance', *Neuroethics*, 15(1): 7. DOI: 10.1007/s12152-022-09481-3

Bublitz, J. C. & Merkel, R. (2009). 'Autonomy and Authenticity of Enhanced Personality Traits', *Bioethics*, 23(6): 360–74. DOI: 10.1111/j.1467-8519.2009.01725.x

Bublitz, J. C. & Merkel, R. (2014). 'Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination', *Criminal Law and Philosophy*, 8(1): 51–77. DOI: 10.1007/s11572-012-9172-y

Buller, T. (2013). 'Neurotechnology, Invasiveness and the Extended Mind', *Neuroethics* , 6(3): 593–605. DOI: 10.1007/s12152-011-9133-5

Buss, S. (2005). 'Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints', *Ethics*, 115(2): 195–235.

Cabrera, L. Y., Evans, E. L., & Hamilton, R. H. (2014). 'Ethics of the Electrified Mind: Defining Issues and Perspectives on the Principled Use of Brain Stimulation in Medical Research and Clinical Care', *Brain Topography* , 27(1): 33–45. DOI: 10.1007/s10548-013-0296-8

Chan, C. K. J. & Harris, J. A. (2017). 'Extinction of Pavlovian Conditioning: The Influence of Trial Number and Reinforcement History', *Behavioural Processes*, 141: 19–25. DOI: 10.1016/j.beproc.2017.04.017

Charter of Fundamental Rights of the European Union (2012). *326/02*.

Chater, N. (2009). 'Rational and Mechanistic Perspectives on Reinforcement Learning', *Cognition*, 113(3): 350–64. DOI: 10.1016/j.cognition.2008.06.014

Chomanski, B. (2023). 'Mental Integrity in the Attention Economy: In Search of the Right to Attention', *Neuroethics*, 16(1): 8. DOI: 10.1007/s12152-022-09514-x

Christman, J. (1991). 'Self-Ownership, Equality, and the Structure of Property Rights', *Political Theory*, 19(1): 28–46. DOI: 10.1177/0090591791019001003

Christman, J. P. (1994). *The Myth of Property: Toward an Egalitarian Theory of Ownership*. Oxford University Press.

Clark, A. & Chalmers, D. (1998). 'The Extended Mind', *Analysis*, 58(1): 7–19. [Analysis Committee, Oxford University Press].

Coates, D. J. & Swenson, P. (2013). 'Reasons-Responsiveness and Degrees of Responsibility', *Philosophical Studies*, 165(2): 629–45. DOI: 10.1007/s11098-012-9969-5

Cohen, F. S. (1954). 'Dialogue on Private Property', *Rutgers Law Review*, 9(2): 357–87.

Cohen, G. A. (1995). *Self-Ownership, Freedom and Equality*. Cambridge University Press.

Cornejo-Plaza, M. I. & Saracini, C. (2023). 'On Pharmacological Neuroenhancement as Part of the New Neurorights' Pioneering Legislation in Chile: A Perspective', *Frontiers in Psychology*, 14: 1177720. DOI: 10.3389/fpsyg.2023.1177720

Craig, J. N. (2016). 'Incarceration, Direct Brain Intervention, and the Right to Mental Integrity—A Reply to Thomas Douglas', *Neuroethics*, 9: 107–18. DOI: 10.1007/s12152-016-9255-x

Cruft, R. (2019). *Human Rights, Ownership, and the Individual. Human Rights, Ownership, and the Individual*. Oxford University Press.

Dekkers, W., Hoffer, C., & Wils, J. P. (2005). 'Bodily Integrity and Male and Female Circumcision', *Medicine, Health Care, and Philosophy*, 8(2): 179–91. DOI: 10.1007/s11019-004-3530-z

De Marco, G. & Douglas, T. (2023). 'Nudge Transparency Is Not Required for Nudge Resistibility', *Ergo*, 10(5): 135–57. DOI: 10.3998/ergo.4635

De Marco, G. & Douglas, T. (2024). 'Do AI Systems Allow Online Advertisers to Control Others?'. Edmonds, D. (ed) *AI Morality*, 1st ed., pp. 150–9. Oxford University Press. DOI: 10.1093/oso/9780198876434.003.0015

De Ridder, D., Kroese, F., & Van Gestel, L. (2022). 'Nudgeability: Mapping Conditions of Susceptibility to Nudge Influence', *Perspectives on Psychological Science*, 17(2): 346–59. DOI: 10.1177/1745691621995183

Dever, J. & Schiller, H. I. (2021). 'This Paper Might Change Your Mind', *Noûs*, 55(4): 863–90. DOI: 10.1111/nous.12336

Do, B., Badillo, M., Cantz, R., & Spivack, J. (2024). 'Privacy and the Rise of "Neurorights" in Latin America'. *Future of Privacy Forum*. Retrieved 3 December 2025 from <https://fpf.org/blog/privacy-and-the-rise-of-neurorights-in-latin-america/#:~:text=As%20a%20pioneer%20in%20the,rule%20on%20a%20neuroprivacy%20case>

Douglas, S. (2011). *Liability for Wrongful Interferences with Chattels*. Hart Publishing.

Douglas, T. (2014). 'Criminal Rehabilitation Through Medical Intervention: Moral Liability and the Right to Bodily Integrity', *The Journal of Ethics*, 18(2): 101–22. DOI: 10.1007/s10892-014-9161-6

Douglas, T. (2025a). 'An Intuitive, Abductive Argument for a Right Against Mental Interference', *The Journal of Ethics*, 29(1): 133–54. DOI: 10.1007/s10892-024-09476-7

Douglas, T. (2025b). 'What Does It Take to Trespass on a Person's Body?', *Analysis*, anaf019. DOI: 10.1093/analys/anaf019

Douglas, T. & Forsberg, L. (2021). 'Three Rationales for a Legal Right to Mental Integrity'. Ligthart, S., van Toor, D., Kooijmans, T., Douglas, T., & Meynen, G. (eds) *Neurolaw*, pp. 179–201. Springer International Publishing. DOI: 10.1007/978-3-030-69277-3_8

Drummond, A. & Sauer, J. D. (2018). 'Video Game Loot Boxes Are Psychologically Akin to Gambling', *Nature Human Behaviour*, 2(8): 530–32. DOI: 10.1038/s41562-018-0360-1

Dworkin, R. (2013). *Taking Rights Seriously*. Bloomsbury Revelations Series. Bloomsbury.

Earp, B. D. (2019). 'The Child's Right to Bodily Integrity'. Edmonds, D. (ed.) *Ethics and the Contemporary World*, pp. 170–85. Routledge.

Eidelson, B. (2020). 'Respect, Individualism, and Colorblindness', *Yale Law Journal*, 129(6): 1600–75.

Engelen, B. (2019). 'Nudging and Rationality: What Is There to Worry?', *Rationality and Society*, 31(2): 204–32. DOI: 10.1177/1043463119846743

Essert, C. (2016). 'Property and Homelessness: Property and Homelessness', *Philosophy & Public Affairs*, 44(4): 266–95. DOI: 10.1111/papa.12080

Faden, R. R., Beauchamp, T. L., & King, N. P. (1986). *A History and Theory of Informed Consent*. Oxford University Press.

Farahany, N. A. (2023). *The Battle for Your Brain: Defending the Right to Think Freely in the Age of Neurotechnology*, 1st ed. [Kindle version]. St. Martin's Press.

Feinberg, J. (1989). 'Personal Sovereignty and Its Boundaries'. *The Moral Limits of the Criminal Law Volume 3: Harm to Self*, pp. 52–97. Oxford University Press.

Feldman, D. (2002). *Civil Liberties and Human Rights in England and Wales*, 2nd ed. Oxford University Press.

Fenwick, M., Siems, M. M., & Wrbka, S. (Eds). (2017). *The Shifting Meaning of Legal Certainty in Comparative and Transnational Law*. Hart Publishing.

Ferster, C. B. & Skinner, B. F. (1957). *Schedules of Reinforcement*. Prentice Hall.

Fins, J. J. (2022). 'The Unintended Consequences of Chile's Neurorights Constitutional Reform: Moving Beyond Negative Rights to Capabilities', *Neuroethics*, 15(3): 26. DOI: 10.1007/s12152-022-09504-z

Fischer, J. M. (2012). 'Responsibility and Autonomy: The Problem of Missions Creep', *Philosophical Issues*, 22(1): 165–84. DOI: 10.1111/j.1533-6077.2012.00223.x

Fischer, J. M. & Ravizza, M. (2000). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge Studies in Philosophy and Law, First Paperback Edition. Cambridge University Press.

Fitz, N. S. & Reiner, P. B. (2015). 'The Challenge of Crafting Policy for Do-It-Yourself Brain Stimulation', *Journal of Medical Ethics* , 41(5): 410–12.

Flanigan, J. (2019). 'Duty and Enforcement', *Journal of Political Philosophy*, 27(3): 341–62. DOI: 10.1111/jopp.12173

Flanigan, J. (2019). 'Boundary Problems and Self-Ownership', *Social Philosophy and Policy*, 36(2): 9–35. DOI: 10.1017/S0265052519000451

Foot, P. (1984). 'Killing and Letting Die'. Garfield, J. L. & Hennessey, P. (eds) *Abortion, Moral and Legal Perspectives*, pp. 177–85. University of Massachusetts Press.

Ford, P. J. & Deshpande, A. (2013). 'The Ethics of Surgically Invasive Neuroscience Research'. Bernat, J. L. & Beresford, H. R. (eds) *Handbook of Clinical Neurology*, Vol. 118, pp. 315–21. Elsevier. DOI: 10.1016/B978-0-444-53501-6.00026-3

Forsberg, L. & Douglas, T. (2016). 'Anti-Libidinal Interventions in Sex Offenders: Medical or Correctional?', *Medical Law Review*, 24(4): 453–73. DOI: 10.1093/medlaw/fww003

Forsberg, L., Douglas, T., & Savulescu, J. (2025). 'Is Consent to Psychological Interventions Less Important than Consent to Bodily Interventions?', *The Philosophical Quarterly*, pqaf005. DOI: 10.1093/pq/pqaf005

Francis, L. P. & Francis, J. G. (1976). 'Nozick's Theory of Rights: A Critical Assessment', *Western Political Quarterly*, 29(4): 634–44. DOI: 10.1177/106591297602900414

Gilbert, D. T. (1991). 'How Mental Systems Believe', *American Psychologist*, 46(2): 107–19.

Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). 'Unbelieving the Unbelievable: Some Problems in the Rejection of False Information', *Journal of Personality and Social Psychology*, 59(4): 601–13. DOI: 10.1037/0022-3514.59.4.601

Gibert, S. (2023). 'The Wrong of Wrongful Manipulation', *Philosophy & Public Affairs*, 51(4): 333–72. DOI: 10.1111/papa.12247

Glannon, W. (2014). 'Neuromodulation, Agency and Autonomy', *Brain Topography*, 27(1): 46–54. DOI: 10.1007/s10548-012-0269-3

Gorin, M. (2014a). 'Do Manipulators Always Threaten Rationality?', *American Philosophical Quarterly*, 51/1: 51–61. [North American Philosophical Publications, University of Illinois Press].

Gorin, M. (2014b). 'Towards a Theory of Interpersonal Manipulation'. Coons, C. & Weber, M. (eds) *Manipulation: Theory and Practice*, pp. 73–97. Oxford University Press. DOI: 10.1093/acprof:oso/9780199338207.003.0004

Government of Spain. (2021). *Charter of Digital Rights*. Retrieved 30 November 2025 from <https://citiesfordigitalrights.org/sites/default/files/140721-Carta_Derechos_Digitales_RedEs_compressed.pdf>

Gregor, M. J. (1996). *Kant: The Metaphysics of Morals*. Cambridge University Press.
Guzmán H. L. (2022). 'Chile: Pioneering the Protection of Neurorights'. *The UNESCO Courier*. Retrieved 30 November 2025 from <https://courier.unesco.org/en/articles/chile-pioneering-protection-neurorights>
Hammer, L. M. (2018). *The International Human Right to Freedom of Conscience: Some Suggestions for Its Development and Application*. Routledge Revivals, Reissued. Routledge, Taylor & Francis Group.
Harris, T. (2016). 'How Technology Is Hijacking Your Mind—From a Former Insider'. *Medium*. Retrieved 24 February 2021 from https://medium.com/thrive-global/how-technology-hijacks-peoples-minds-from-a-magician-and-google-s-design-ethicist-56d62ef5edf3
Hausman, D. M. & Welch, B. (2010). 'Debate: To Nudge or Not to Nudge', *Journal of Political Philosophy*, 18(1): 123–36.
Heering, D. (2022). 'Actual Sequences, Frankfurt-Cases, and Non-Accidentality', *Inquiry*, 65(10): 1269–88. DOI: 10.1080/0020174X.2021.1904644
Herring, J. & Wall, J. (2017). 'The Nature and Significance of the Right to Bodily Integrity', *The Cambridge Law Journal*, 76(03): 566–88. DOI: 10.1017/S0008197317000605
Hertz, N. (2023). 'Neurorights—Do We Need New Human Rights? A Reconsideration of the Right to Freedom of Thought', *Neuroethics*, 16(1): 5. DOI: 10.1007/s12152-022-09511-0
Holm, S. (2025). 'A Right to Bodily Integrity—Some Complications', *Clinical Ethics*, 20(2): 69–72. DOI: 10.1177/14777509251321907
Honoré, A. M. (1961). 'Ownership'. Guest, A. G. (ed.) *Oxford Essays in Jurisprudence*, pp. 370–75. Clarendon Press.
Ienca, M. & Andorno, R. (2017). 'Towards New Human Rights in the Age of Neuroscience and Neurotechnology', *Life Sciences, Society and Policy*, 13: 5. DOI: 10.1186/s40504-017-0050-1
Ienca, M. & Andorno, R. (2017). 'A New Category of Human Rights: Neurorights'. *SpringerOpen Blog*. Retrieved from <http://blogs.springeropen.com/springeropen/2017/04/26/new-category-human-rights-neurorights/>
Ivanković, V. & Engelen, B. (2019). 'Nudging, Transparency, and Watchfulness', *Social Theory and Practice*, 45(1): 43–73.
James, S. (2003). 'VII-Rights as Enforceable Claims', *Proceedings of the Aristotelian Society*, 103(1): 133–47. DOI: 10.1111/j.0066-7372.2003.00067.x
Jenkin, Z. (2023). 'Perceptual Learning and Reasons-Responsiveness', *Noûs*, 57(2): 481–508. DOI: 10.1111/nous.12425
Johansson, J. & Risberg, O. (2019). 'The Preemption Problem', *Philosophical Studies* , 176(2): 351–65.
Kaiserman, A. (2021). 'Reasons-Sensitivity and Degrees of Free Will', *Philosophy and Phenomenological Research*, 103(3): 687–709. DOI: 10.1111/phpr.12738
Kamm, F. M. (1992). 'Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status', *Philosophy & Public Affairs*, 21(4): 354–89.
Kamm, F. M. (2004). 'Rights'. Coleman, J. L., Shapiro, S., & Himma, K. E. (eds) *The Oxford Handbook of Jurisprudence and Philosophy of Law*, pp. 476–513. Oxford University Press.
Kant, I. (1963). *Lectures on Ethics* (L. Infield, Trans.) Library of Religion and Culture. Harper & Row.
Karremans, J. C., Stroebe, W., & Claus, J. (2006). 'Beyond Vicary's Fantasies: The Impact of Subliminal Priming and Brand Choice', *Journal of Experimental Social Psychology*, 42(6): 792–98. DOI: 10.1016/j.jesp.2005.12.002
Kipper, J. (2021). 'Irresistible Nudges, Inevitable Nudges, and the Freedom to Choose', *Moral Philosophy and Politics*, 8(2): 285–303. De Gruyter. DOI: 10.1515/mopp-2020-0013

Klenk, M. (2021). 'Interpersonal Manipulation', *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3859178

Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199552795.001.0001

Kroese, F. M., Marchiori, D. R., & de Ridder, D. T. D. (2016). 'Nudging Healthy Food Choices: A Field Experiment at the Train Station', *Journal of Public Health*, 38(2): e133–37. DOI: 10.1093/pubmed/fdv096

Lavazza, A. & Giorgi, R. (2023). 'Philosophical Foundation of the Right to Mental Integrity in the Age of Neurotechnologies', *Neuroethics*, 16(1): 10. DOI: 10.1007/s12152-023-09517-2

Lawrence, N. (2017). 'The Troubling Psychology of Pay-to-Loot Systems'. *IGN*. Retrieved 20 December 2022 from <https://www.ign.com/articles/2017/04/24/the-troubling-psychology-of-pay-to-loot-systems>

Levy, N. (2007). *Neuroethics: Challenges for the 21st Century*. Cambridge University Press.

Levy, N. (2007). 'Rethinking Neuroethics in the Light of the Extended Mind Thesis', *The American Journal of Bioethics*, 7(9): 3–11. DOI: 10.1080/15265160701518466

Levy, N. (2019). 'Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons', *Ergo*, 6(10): 281–302. DOI: http://dx.doi.org/10.3998/ergo.12405314.0006.010

Ligthart, S., Bublitz, C., & Alegre, S. (2023). 'Neurotechnology: We Need New Laws, Not New Rights', *Nature*, 620(7976): 950. DOI: 10.1038/d41586-023-02698-z

Ligthart, S., Ienca, M., Meynen, G., Molnar-Gabor, F., Andorno, R., Bublitz, C., Catley, P., et al. (2023). 'Minding Rights: Mapping Ethical and Legal Foundations of "Neurorights"', *Cambridge Quarterly of Healthcare Ethics*, 32(4): 461–81. DOI: 10.1017/S0963180123000245

Lippert-Rasmussen, K. (2008). 'Against Self-Ownership: There Are No Fact-Insensitive Ownership Rights over One's Body', *Philosophy & Public Affairs*, 36(1): 86–118. DOI: 10.1111/j.1088-4963.2008.00125.x

Lippert-Rasmussen, K. (2018). 'The Self-Ownership Trilemma, Extended Minds, and Neurointerventions'. Birks, D. & Douglas, T. (eds) *Treatment for Crime: Philosophical Essays on Neurointerventions in Criminal Justice*, pp. 140–58. Oxford University Press.

Locke, J. (1849). *An Essay Concerning Human Understanding*, 30th ed., carefully rev. and compared with the best copies. W. Tegg.

Locke, J. (1982). *Second Treatise of Government: An Essay Concerning the True Original, Extent and End of Civil Government* (R. H. Cox, Ed.). John Wiley & Sons, Incorporated.

López-Silva, P. (Ed). (2025). *Contextualizing Neuroprotection: Latin American Perspectives on the Impact of Neurotechnological Development in Life and Society. Ethics of Science and Technology Assessment,* 1st ed. Springer Nature Switzerland. DOI: 10.1007/978-3-031-96055-0

MacKinnon, C. A. (1993). *Only Words*. Harvard University Press.

Maestri v Italy. (2004), Vol. Series A.

Mandelbaum, E. (2014). 'Thinking is Believing', *Inquiry*, 57(1): 55–96. DOI: 10.1080/0020174X.2014.858417

McCarthy-Jones, S. (2021). 'Freedom of Thought: Who, What, and Why?'. Blitz, M. J. & Bublitz, C. (eds) *The Law and Ethics of Freedom of Thought*, Vol. 1, pp. 27–47. Palgrave Macmillan.

McCarthy-Jones, S. (2019). 'The Autonomous Mind: The Right to Freedom of Thought in the Twenty-First Century', *Frontiers in Artificial Intelligence*, 2: 19. DOI: 10.3389/frai.2019.00019

McCay, A. (2022). 'Neurorights: The Chilean Constitutional Change', *AI & SOCIETY*, 39: 797–98. DOI: 10.1007/s00146-022-01396-0

McDermott, M. (1982). 'Utility and Population', *Philosophical Studies*, 42(2): 163–77. DOI: 10.1007/BF00374031

McMahan, J. (2002). *The Ethics of Killing: Problems at the Margins of Life*. Oxford University Press.

McKenzie, C. R. M., Liersch, M. J., & Finkelstein, S. R. (2006). 'Recommendations Implicit in Policy Defaults', *Psychological Science*, 17(5): 414–20.

Mele, A. R. (2016). 'Moral Responsibility: Radical Reversals and Original Designs', *The Journal of Ethics*, 20(1–3): 69–82. Springer Verlag.

Mele, A. R. (2001). *Autonomous Agents: From Self-Control to Autonomy*, issued as an Oxford University Paperback. Oxford University Press.

Mendlow, G. S. (2018). 'Why Is It Wrong to Punish Thought?', *Yale Law Journal*, 127(8): 2342–86. DOI: 10.1007/978-3-030-84494-3_5

Mills, C. (2015). 'The Heteronomy of Choice Architecture', *Review of Philosophy and Psychology*, 6(3): 495–509. DOI: 10.1007/s13164-015-0242-7

Mill, J. S. (1909). *On Liberty*. The Floating Press.

Montminy, M. & Tinney, D. (2018). 'Manipulation and Degrees of Blameworthiness', *The Journal of Ethics*, 22(3–4): 265–81. DOI: 10.1007/s10892-018-9274-4

Münscher, R., Vetter, M., & Scheuerle, T. (2016). 'A Review and Taxonomy of Choice Architecture Techniques', *Journal of Behavioral Decision Making*, 29(5): 511–24. DOI: 10.1002/bdm.1897

Murdoch, J. (2007). *Freedom of Thought, Conscience and Religion*. Directorate General of Human Rights and Legal Affairs, Council of Europe. Retrieved 30 November 2025 from <https://rm.coe.int/168007ff4f>

Nagel, T. (2004). 'Personal Rights and Public Space'. *Concealment and Exposure: And Other Essays*, pp. 31–52. Oxford University Press.

Nielsen, R. K. L. & Grabarczyk, P. (2019). 'Are Loot Boxes Gambling? Random Reward Mechanisms in Video Games', *Transactions of the Digital Games Research Association*, 4(3): 171–207. DOI: 10.26503/todigra.v4i3.104

Noggle, R. (2020). 'Pressure, Trickery, and a Unified Account of Manipulation', *American Philosophical Quarterly*, 3(57): 241–52.

Noggle, R. (2025). *Manipulation: Its Nature, Mechanisms, and Moral Status*. New Topics in Applied Philosophy, 1st ed. Oxford University Press.

Nozick, R. (1974). *Anarchy, State and Utopia*. Basic Books.

Olson, E. T. (2003). 'An Argument for Animalism'. Martin, R. & Barresi, J. (eds) *Personal Identity*, pp. 318–34. Blackwell Readings in Philosophy. Blackwell

Otsuka, M. (1998). 'Self-Ownership and Equality: A Lockean Reconciliation', *Philosophy & Public Affairs*, 27(1): 65–92. DOI: 10.1111/j.1088-4963.1998.tb00061.x

Owens, D. (2019). 'Property and Authority', *Journal of Political Philosophy*, 27(3): 271–93. DOI: 10.1111/jopp.12190

Pallikkathayil, J. (2017). 'Persons and Bodies'. Kisilevsky, S. & Stone, M. J. (eds) *Freedom and Force: Essays on Kant's Legal Philosophy*, pp. 35–54. Hart Publishing. DOI: 10.5040/9781474201889

Parekh, B. (2012). 'Is There a Case for Banning Hate Speech?'. Herz, M. & Molnar, P. (eds) *The Content and Context of Hate Speech*, 1st ed., pp. 37–56. Cambridge University Press. DOI: 10.1017/CBO9781139042871.006

Penner, J. (2000). *The Idea of Property in Law*. *The Idea of Property in Law*. Oxford University Press.

Pereboom, D. (2016). *Free Will, Agency, and Meaning in Life*. Oxford University Press.

Plassmann, H., Ramsøy, T. Z., & Milosavljevic, M. (2012). 'Branding the Brain: A Critical Review and Outlook', *Journal of Consumer Psychology*, 22(1): 18–36. DOI: 10.1016/j.jcps.2011.11.010

Plato. (2019). *Gorgias* (R. Waterfield, ed.) Oxford World's Classics. Oxford University Press.
Pugh, J. (2020). *Autonomy, Rationality, and Contemporary Bioethics*. Oxford Philosophical Monographs, 1st ed. Oxford University Press.
Puri, A. (2021). 'The Right to Attentional Privacy', *Rutgers Law Record*, 48(1): 206–21.
Quilty-Dunn, J. & Mandelbaum, E. (2018). 'Inferential Transitions', *Australasian Journal of Philosophy*, 96(3): 532–47. DOI: 10.1080/00048402.2017.1358754
Quinn, W. S. (1989). 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *Philosophical Review* , 98(3): 287–312. Duke University Press.
Quong, J. (2011). *Liberalism Without Perfection*. Oxford University Press.
R (on the application of Justin West) v The Parole Board. (2002). *EWCA Civ*, Vol. [2002].
Ramnerö, J., Molander, O., Lindner, P., & Carlbring, P. (2019). 'What Can Be Learned About Gambling from a Learning Perspective? A Narrative Review', *Nordic Psychology*, 71(4): 303–22. DOI: 10.1080/19012276.2019.1616320
Ratoff, W. (2024). 'The Right to Mental Autonomy: Its Nature and Scope', *Journal of Ethics and Social Philosophy*, 27(2): 257–86.
Raz, J. (1986). *The Morality of Freedom*. Clarendon Press; Oxford University Press.
Ripstein, A. (2006). 'Beyond the Harm Principle', *Philosophy & Public Affairs* , 34(3): 215–45.
Roskies, A. (2016). 'Neuroethics'. Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*, Spring 2016. Metaphysics Research Lab, Stanford University.
Russell, D. C. (2010). 'Embodiment and Self-Ownership', *Social Philosophy and Policy*, 27(1): 135–67. Cambridge University Press. DOI: 10.1017/S0265052509990069
Ryberg, J. (2019). *Neurointerventions, Crime, and Punishment: Ethical Considerations*., 1st ed. Oxford University Press. DOI: 10.1093/oso/9780190846428.001.0001
Ryberg, J. (2020). *Neurointerventions, Crime, and Punishment: Ethical Considerations*. Studies in Penal Theory and Philosophy. Oxford University Press.
Saghai, Y. (2013). 'Salvaging the Concept of Nudge', *Journal of Medical Ethics*, 39(8): 487–93. DOI: 10.1136/medethics-2012-100727
Sartorio, C. (2016). *Causation and Free Will*. Oxford University Press.
Sartorio, C. (2018). 'Situations and Responsiveness to Reasons: Situations and Responsiveness to Reasons', *Noûs*, 52(4): 796–807. DOI: 10.1111/nous.12212
Schauer, F. (2020). 'Freedom of Thought?', *Social Philosophy and Policy*, 37(2): 72–89. DOI: 10.1017/S0265052521000054
Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press. DOI: 10.1017/CBO9781139173438
Searle, J. R. (1976). 'A Classification of Illocutionary Acts', *Language in Society*, 5(1): 1–23. DOI: 10.1017/S0047404500006837
Shaw, E. (2022). 'Neuroscience, Criminal Sentencing, and Human Rights', *William & Mary Law Review*, 63: 1409–43.
Simons, D. J. & Chabris, C. F. (2010). *The Invisible Gorilla: And Other Ways Our Intuitions Deceive Us*. Crown Publishing Group.
Solon, O. (2017). 'Ex-Facebook President Sean Parker: Site Made to Exploit Human "Vulnerability"'. *The Guardian*. Retrieved 30 November 2025 from <http://www.theguardian.com/technology/2017/nov/09/facebook-sean-parker-vulnerability-brain-psychology>
Steffel, M., Williams, E. F., & Pogacar, R. (2016). 'Ethically Deployed Defaults: Transparency and Consumer Protection through Disclosure and Preference Articulation', *Journal of Marketing Research*, 53(5): 865–80. DOI: 10.1509/jmr.14.0421
Sunstein, C. (2014). *Why Nudge?: The Politics of Libertarian Paternalism*, pp. 1–195. Yale University Press.

Susser, D., Roessler, B., & Nissenbaum, H. (2019). 'Technology, Autonomy, and Manipulation', *Internet Policy Review*, 8(2): 1–22. DOI: 10.14763/2019.2.1410

Taube, O. & Vetter, M. (2019). 'How Green Defaults Promote Environmentally Friendly Decisions: Attitude-Conditional Default Acceptance But Attitude-Unconditional Effects on Actual Choices', *Journal of Applied Social Psychology*, 49(11): 721–32. DOI: 10.1111/jasp.12629

Taylor, J. S. (2009). *Practical Autonomy and Bioethics*. Routledge. DOI: 10.4324/9780203873991

Taylor, R. S. (2004). 'A Kantian Defense of Self-Ownership', *Journal of Political Philosophy*, 12(1): 65–78. DOI: 10.1111/j.1467-9760.2004.00191.x

Tesink, V., Douglas, T., Forsberg, L., Ligthart, S., & Meynen, G. (2023). 'Neurointerventions in Criminal Justice: On the Scope of the Moral Right to Bodily Integrity', *Neuroethics* , 16(3): 26. DOI: 10.1007/s12152-023-09526-1

Thaler, R. H. & Sunstein, C. R. (2008). *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Yale University Press.

Thibaut, F., Barra, F. D. L., Gordon, H., Cosyns, P., & Bradford, J. M. W. (2010). 'The World Federation of Societies of Biological Psychiatry (WFSBP) Guidelines for the Biological Treatment of Paraphilias', *World Journal of Biological Psychiatry*, 11(4): 604–55.

Thomson, J. J. (1990). 'Trespass and First Property'. *The Realm of Rights*, pp. 205–26. Harvard University Press.

Todini, L. & Fantuz, F. (2023). 'Thirst: Neuroendocrine Regulation in Mammals', *Veterinary Research Communications*, 47(3): 1085–101. DOI: 10.1007/s11259-023-10104-2

Trachsel, M., Holtforth, M. G., Biller-Andorno, N., & Appelbaum, P. S. (2015). 'Informed Consent for Psychotherapy: Still Not Routine', *The Lancet Psychiatry*, 2(9): 775–77. DOI: 10.1016/S2215-0366(15)00318-1

Tran, J. L. (2016). 'The Right to Attention', *Indiana Law Journal*, 91(3): 1023–62.

Trump, C. E., Herrod, J. L., Ayres, K. M., Ringdahl, J. E., & Best, L. (2021). 'Behavior Momentum Theory and Humans: A Review of the Literature', *The Psychological Record*, 71(1): 71–83. DOI: 10.1007/s40732-020-00430-1

UN General Assembly. (1948). 'Third Committee, Third Session, Meeting 127'. Paris, France.

Vallentyne, P. (1998). 'Critical Notice of G.A. Cohen's Self-Ownership, Freedom, and Equality', *Canadian Journal of Philosophy*, 28: 609–26.

Vallentyne, P. (2011). 'Nozick's Libertarian Theory of Justice'. Bader, R. & Meadowcroft, J. (eds) *The Cambridge Companion to Nozick's Anarchy, State, and Utopia*, pp. 145–67. Cambridge University Press.

Véliz, C. (2020). *Privacy Is Power*. Penguin (Bantam Press).

Veltkamp, M., Custers, R., & Aarts, H. (2011). 'Motivating Consumer Behavior by Subliminal Conditioning in the Absence of Basic Needs: Striking Even While the Iron Is Cold', *Journal of Consumer Psychology*, 21(1): 49–56. DOI: 10.1016/j.jcps.2010.09.011

Vermeulen, B. & van Roosmalen, M. (2018). 'Freedom of Thought, Conscience and Religion (Article 9)'. van Dijk, P., van Hoof, G. J. H., van Rijn, A. B., & Zwaak, L. (eds) *Theory and Practice of the European Convention on Human Rights*, pp. 751–77. Cambridge.

Verwijmeren, T., Karremans, J. C., Stroebe, W., & Wigboldus, D. H. J. (2011). 'The Workings and Limits of Subliminal Advertising: The Role of Habits', *Journal of Consumer Psychology*, 21(2): 206–13. DOI: 10.1016/j.jcps.2010.11.004

Viens, A. M. (2020). 'The Right to Bodily Integrity'. *Cambridge Handbook of New Human Rights*, pp. 361–84. Cambridge University Press.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oxford University Press.

Waldron, J. (1990). 'Locke's Discussion of Property'. *The Right to Private Property*, pp. 165–206. Oxford University Press.

Waldron, J. (2014). *The Harm in Hate Speech*., First Harvard University Press Paperback Edition. Harvard University Press.

Weinrib, E. J. (1995). *The Idea of Private Law*. Harvard University Press.

Wertheimer, A. (2003). 'The Harm and Wrong of Rape'. *Consent to Sexual Relations*, Cambridge Studies in Philosophy and Law, pp. 89–118. Cambridge University Press. DOI: 10.1017/CBO9780511610011.006

Wilkinson, T. M. (2011). *Ethics and the Acquisition of Organs*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199607860.001.0001

Wilkinson, T. M. (2013). 'Nudging and Manipulation', *Political Studies*, 61(2): 341–55. DOI: 10.1111/j.1467-9248.2012.00974.x

Wong, J. C. (2017). 'Former Facebook Executive: Social Media Is Ripping Society Apart'. *The Guardian*. Retrieved 30 November 2025 from <http://www.theguardian.com/technology/2017/dec/11/facebook-former-executive-ripping-society-apart>

Wood, A. W. (2014). 'Coercion, Manipulation, Exploitation'. Coons, C. & Weber, M. (eds) *Manipulation: Theory and Practice*, pp. 17–51. Oxford University Press.

Zigmond, T. (2017). 'Mental Health Law Across the UK', *BJPsych Bulletin*, 41(6): 305–7. Cambridge University Press.

Zuboff, S. (2020). 'You Are Now Remotely Controlled'. *The New York Times*. Retrieved 30 November 2025 from <https://www.nytimes.com/2020/01/24/opinion/sunday/surveillance-capitalism.html>

Index

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.