

OXFORD SERIES IN NEUROSCIENCE, LAW,
AND PHILOSOPHY

Series Editors

Lynn Nadel, Frederick Schauer, and Walter P. Sinnott-Armstrong

Conscious Will and Responsibility

Edited by Walter P. Sinnott-Armstrong and Lynn Nadel

Memory and Law

Edited by Lynn Nadel and Walter P. Sinnott-Armstrong

Neuroscience and Legal Responsibility

Edited by Nicole A Vincent

Handbook on Psychopathy and Law

Edited by Kent A. Kiehl and Walter P. Sinnott-Armstrong

A Primer on Criminal Law and Neuroscience

Edited by Stephen J. Morse and Adina L. Roskies

Bioprediction, Biomarkers, and Bad Behavior

Edited by Ilina Singh, Walter P. Sinnott-Armstrong, and Julian Savulescu

Finding Consciousness: The Neuroscience, Ethics, and Law of Severe Brain Damage

Edited by Walter Sinnott-Armstrong

*The Neuroethics of Biomarkers: What the Development of Bioprediction Means for
Moral Responsibility, Justice, and the Nature of Mental Disorder*

By Matthew L. Baum

Neurointerventions and the Law: Regulating Human Mental Capacity

Edited by Nicole A Vincent, Thomas Nadelhoffer, and Allan McCay

Neurointerventions and the Law

Regulating Human Mental Capacity

Edited by

NICOLE A VINCENT, THOMAS NADELHOFFER,
AND ALLAN MCCAY

OXFORD
UNIVERSITY PRESS

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 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.

© Oxford University Press 2020

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form
and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data

Names: Vincent, Nicole A, editor. | Nadelhoffer, Thomas, editor. | McCay, Allan, editor.

Title: Neurointerventions and the law : regulating human mental capacity /

Nicole A Vincent, Thomas Nadelhoffer, Allan McCay (editors).

Description: New York, NY : Oxford University Press, 2020. |

Series: Oxford series in neuroscience, law, and philosophy |

Includes bibliographical references and index.

Identifiers: 2019048935 (print) | 2019048936 (ebook) | ISBN 9780190651145 (hardback) |

ISBN 9780190667979 (epub) | ISBN 9780190651169

Subjects: LCSH: Mental health laws. | Law—Psychological aspects. |

Criminal psychology. | Neurology—Law and legislation

Classification: LCC K3608 .N48 2020 (print) | LCC K3608 (ebook) | DDC 344.04/4—dc23

LC record available at <https://lccn.loc.gov/2019048935>

LC ebook record available at <https://lccn.loc.gov/2019048936>

1 3 5 7 9 8 6 4 2

Printed by Sheridan Books, Inc., United States of America

<i>Contributors</i>	vii
1. Law Viewed Through the Lens of Neurointerventions <i>Nicole A Vincent, Thomas Nadelhoffer, and Allan McCay</i>	1
PART I. CONCEPTUAL, ETHICAL, AND JURISPRUDENTIAL ISSUES	
2. Cognitive Enhancement: Defending the Parity Principle <i>Neil Levy</i>	33
3. Why Means Matter: Legally Relevant Differences Between Direct and Indirect Interventions into Other Minds <i>Jan Christoph Bublitz</i>	49
4. Neuroprosthetics, Behavior Control, and Criminal Responsibility <i>Walter Glannon</i>	89
5. Is There Anything Wrong With Using AI Implantable Brain Devices to Prevent Convicted Offenders from Reoffending? <i>Frédéric Gilbert and Susan Dodds</i>	113
6. Offering Neurointerventions to Offenders With Cognitive- Emotional Impairments: Ethical and Criminal Justice Aspects <i>Farah Focquaert, Kristof Van Assche, and Sigrid Sterckx</i>	127
7. Diversion Courts, Traumatic Brain Injury, and American Vets <i>Valerie Gray Hardcastle</i>	149
8. Neurobionic Revenge Porn and the Criminal Law: Brain-Computer Interfaces and Intimate Image Abuse <i>Allan McCay</i>	168
PART II. PUNISHING PEOPLE	
9. Folk Jurisprudence and Neurointervention: An Interdisciplinary Investigation <i>Thomas Nadelhoffer, Daniela Goya-Tocchetto, Jennifer Cole Wright, and Quinn McGuire</i>	191

10. Judicious Use of Neuropsychiatric Evidence When Sentencing Offenders With Addictive Behaviors: Implications for Neurointerventions 231
Andrew Dawson, Jennifer Chandler, Colin Gavaghan, Wayne Hall, and Adrian Carter

PART III. HEALING PEOPLE

11. "It Will Help You Repent": Why the Communicative Theory of Punishment Requires the Provision of Medications to Offenders With ADHD 255
William Bülow
12. Is It *Really* Ethical to Prescribe Antiandrogens to Sex Offenders to Decrease Their Risk of Recidivism? 270
Christopher James Ryan
13. Chemical Castration as Punishment 293
Katrina L. Sifferd
14. Foundational Facts for Legal Responsibility: Human Agency and the Aims of Restorative Neurointerventions 319
Paul Sheldon Davies

PART IV. CHANGING PEOPLE

15. Make Me Gay: What Neurointerventions Tell Us About Sexual Orientation and Why It Matters for the Law 351
Andrew Vierra

PART V. ENHANCING PEOPLE

16. Neuroenhancement, Coercion, and Neo-Luddism 375
Alexandre Erler
17. Neurointerventions and Business Law: On the Legal and Moral Issues of Neurotechnology in Business and How They Differ From the Criminal Law Context 406
Patrick D. Hopkins and Harvey L. Fiser

- Index* 433

Contributors

Jan Christoph Bublitz, LLB, JD, PhD
Postdoctoral Researcher
University of Hamburg
Hamburg, Germany

William Bülow, PhD
Postdoctoral Researcher
Stockholm Centre for the Ethics of
War and Peace
Stockholm University
Stockholm, Sweden

Adrian Carter, PhD
Associate Professor
School of Psychological Sciences
Faculty of Medicine Nursing and
Health Sciences
Monash University
Victoria, Australia

Jennifer Chandler, LLM, LLB
Professor of Law
University of Ottawa
Ottawa
Canada, ON

Paul Sheldon Davies, PhD
Professor of Philosophy
College of William & Mary
Williamsburg, VA

Andrew Dawson, PhD
Strategic Project Adviser
(Community & Government)
Turner Institute for Brain and
Mental Health Medicine,
Nursing and Health Sciences
Monash University
Victoria, Australia

Susan Dodds, PhD
Deputy Vice-Chancellor and
Vice-President
Research and Industry
Engagement
La Trobe University
Victoria, Australia

Alexandre Erler, DPhil
Research Assistant Professor
Philosophy and Bioethics
Chinese University
of Hong Kong
Hong Kong, SAR

Harvey L. Fiser, JD
Associate Professor of Law
Elsë School of Management
Millsaps University
Jackson, MS

Farah Focquaert, PhD
Professor
Philosophical Anthropology
Ghent University
Ghent, Belgium

Colin Gavaghan, LLB, PhD
Associate Professor
Law and Policy
University of Otago
Dunedin, New Zealand

Frédéric Gilbert, PhD
Senior Lecturer in Philosophy
Philosophy and Gender Studies
University of Tasmania
Hobart, Australia

- Noble, K. G., McCandliss, B. D., & Farah, M. J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, 10, 464–480.
- Randall, D. C., Shneerson, J. M., & File, S. E. (2005). Cognitive effects of modafinil in student volunteers may depend on IQ. *Pharmacology Biochemistry and Behavior*, 82, 133–139.
- Sahakian, B. J., & Morein-Zamir, S. (2007). Professor's little helper. *Nature*, 450, 1157–1159.
- Schermer, M. (2008). On the argument that enhancement is “cheating.” *Journal of Medical Ethics*, 34, 85–88.
- Turner, D. C., Clark, L., Dowson, J., Robbins, T. W., & Sahakian, B. J. (2004). Modafinil improves cognition and response inhibition in adult ADHD. *Biological Psychiatry*, 5, 1031–1040.
- Uhlmann, E. L., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, 16, 474–480.
- Wansink, B., Kent, R. J., & Hoch, S. J. (1998). An anchoring and adjustment model of purchase quantity decisions. *Journal of Marketing Research*, 35, 71–81.

Why Means Matter

Legally Relevant Differences Between Direct and Indirect Interventions into Other Minds

Jan Christoph Bublitz

There is nothing in the mind that has not been in the senses—except the mind itself.

—Gottfried Wilhelm Leibniz

Introduction

Are direct interventions into brains and minds, especially novel neuro-technological ones, inherently different to indirect ways of changing minds? This is a key question of neuroethics that any legal regulation of mind-interventions—old or new, natural or technological—has to face.¹ I wish to provide an affirmative answer supportive of such differences. This requires a twofold argument. It has to show, first, that there are differences between direct (or synonymously, biological or physiological) and indirect (psychological) interventions which are not based on crude mind–brain dualisms or dubious properties such as the naturalness of an intervention. Second, it has to demonstrate why these differences (should) matter for the law. This is the program for this chapter. In a nutshell, I propose understanding indirect interventions as stimuli that persons perceive through their external senses and direct interventions as those that reach brains and minds on different, nonperceptual routes. Interventions thus primarily differ in virtue of their causal pathways, because of which persons have different kinds and amounts of control over interventions. Direct interventions change minds by *bypassing resistance and control* of recipients, quite unlike mind-changes caused by perceptual inputs. In addition, direct interventions differ

¹ This paper clarifies and expands upon previous work co-authored with Reinhard Merkel (Bublitz & Merkel, 2014). It greatly benefited from critical remarks especially by Neil Levy, Thomas Douglas, Nicole Vincent, and Allan McCay. I wish to thank all of them and address some of their points in due course. The epigraph by Leibniz (1709) commemorates the 300th anniversary of his death at the time of this writing.

from indirect ones in the way they relate to the ordinary functions of the mind-brain system, if applied without consent, they *misappropriate* mechanisms of the brain. These differences bear normative relevance in light of what I suggest to be the guiding normative principle in this domain, the human right to mental self-determination (or cognitive liberty). As a consequence, I propose the law should adopt a rough normative—not ontological—dualism between interventions into other minds: nonconsensual direct interventions into other minds should be prohibited, with few exceptions. By contrast, indirect interventions should be *prima facie* permissible, primarily those that qualify as exercises of free speech or other protected rights of interveners. Nonetheless, indirect interventions may require further context specific evaluations, and some may flout mental self-determination of targets to an extent that indicates a need for their restriction (e.g., subliminal stimuli).

Put differently, I forward the claim that assessments of direct and indirect interventions touch upon relevant normative considerations to different degrees. Typical justifications for direct and indirect interventions run differently, indirect ones usually fare better in light of applicable norms. Therefore, treating direct and indirect interventions on a par, as famously suggested by Neil Levy (2007, defended in chapter 2 in this volume), is neither normatively warranted, nor heuristically helpful, at least in more fine-grained evaluations of neurointerventions which legal regulations require.

To set the stage, it is worthy to note that people change each other's minds all the time. Humans are social cognizers, naturally reading and influencing other minds. Some consider the capacity to shape minds as the linchpin of the human cognitive system (Zawidzki, 2013). At least, it is a key feature that sets it apart from those of other species (Tomasselo, 2016). Accordingly, humans regularly alter beliefs, elicit emotions, or modulate various mental states and processes in others.² Sometimes people change others' minds intentionally, sometimes through trickery, and sometimes unexpectedly or accidentally. Regularly, people alter others' minds through words and at times, through more sophisticated means from psychotherapy to pharmaceuticals or novel neurotechnological devices. The primary aim of such interventions is altering some aspect of the mental world of addressees by producing rather specific or more diffuse mental effects. The question is whether particular means—or classes of means—to intervene into minds are normatively more problematic than others. Public opinion commonly draws a distinction between direct and indirect interventions whereby the former are conceived as more problematic than the

latter. Direct interventions comprise psychotropic drugs (pharmaceuticals), psychosurgery, electric or magnetic brain stimulation (through techniques such as transcranial direct current stimulation [tDCS], transcranial magnetic stimulation [TMS], deep brain stimulation [DBS]),³ and, less familiar, methods such as ultrasound (Martin, Jeanmonod, Morel, Zadicario, & Werner, 2009) and possibly optogenetic tools in the future (Anderrson et al., 2016). These biological or physiological interventions are often contrasted with supposedly less unsettling indirect interventions, from verbal communication over psychotherapy to visual stimuli or music.⁴ However, especially the pioneering work of Levy has cast doubt on the plausibility of distinctions between direct and indirect interventions and their ethical relevance. He argues that many of the criteria for drawing distinctions between interventions are misleading for empirical, metaphysical or normative reasons. Consequently, he claims that the nature of an intervention (direct–indirect, traditional–novel, natural–artificial) is irrelevant for its ethical assessment. Only the *effects* of interventions matter. Accordingly, he proposes that different (classes of) interventions should be treated on a par as long as their effects are relevantly similar. Levy's so-called parity principle strikes a chord because it rightfully exposes concerns with neurobiological interventions, which often remain vague and seem to be driven by bioconservative sentiments. However, although Levy forcefully shows that many worries over direct interventions are ill-founded, I wish to claim that not *all* of them are. The following seeks to stake out a third position, situated between the parity principle and bioconservative rejections of allegedly antinatural interventions. I will thus vindicate the public skepticism about direct intervention, but on different grounds. Normatively, my position is based in the idea of mental or psychological self-determination, and formally in an understanding of mind-interventions as social interactions between different persons with legitimate interests on either side. Therefore—and this might be a more general methodological point—the formal (legal) relations between interveners and addressees, or senders and receivers, with rights and interests on either side have to be recognized. Arguments and evaluations that dwell only one-sidedly on effects on recipients and neglect relational considerations and rights of senders miss important aspects and are prone to draw misleading analogies or generalizations. For instance, different norms may apply in evaluations of what

³ A comprehensive recent introduction to the varieties of brain stimulation methods is Reti (2015). For its dark history, see Valenstein (1973).

⁴ This asymmetry permeates public debates over therapeutic or enhancing drugs. Its appeal is confirmed by a recent study by Specker, Schermer, and Reiner (2017). The distinction also partially mirrors the disciplinary divide between psychology and psychiatry (and their respective treatment modalities). Traces of the distinction can also be found in the law. One example is a dualism found in criminal law which often provides strong protection to bodies against harmful interventions, but only fragmentarily to minds (Bublitz & Merkel, 2014).

² For brevity's sake, I speak of mental states, but this includes all kinds of mental events or processes, conscious or nonconscious. Also, *recipients*, *addressees*, and *targets* of interventions are used interchangeably.

people do to themselves, what parents do to their children, or the state does to criminal offenders. Accordingly, direct interventions might be less problematic in one case and more in another. Moreover, indirect interventions regularly fare better because some rights entitle interveners to alter other minds indirectly, most notably the right to freedom of speech, whereas there are usually no rights that entitle people to directly intervene into other minds. In other words, direct interventions usually do not manifest legally protected interests, but indirect ones do. The divergence of rights on either side is a key difference between direct and indirect interventions and flows from my proposed understanding of mental self-determination. Appreciating these dissimilarities undermines the parity principle and opens the view for more nuanced evaluations of particular interventions. Levy raises some objections to (previous formulations of) my account in the previous chapter to which I respond in the final section.

At the outset, it is helpful to contextualize the dispute. In the long run, legal scholars and policymakers are interested in assessing whether the use of particular means for particular ends in specific contexts is morally or legally (im)permissible. The law has to develop respective norms and doctrines that clearly convey to potential interveners which means of changing minds are permissible and which are not. Correspondingly, the law has to define whether and under which conditions affected persons can legitimately complain about—or even forcefully resist—unwanted alterations of their minds and the conditions under which the state is obliged to provide protection against unwanted interventions. Answering these questions requires developing a legal taxonomy of permissible and impermissible mind-interventions. It has to accommodate a range of broad as well as context-specific considerations. The distinction between direct and indirect interventions is just one of them. Therefore, even if the normative dualism suggested here holds, it may still be the case that, all things considered, some direct interventions are unproblematic in some instances whereas indirect ones are objectionable in others. What is legitimate in romantic seduction might, for instance, turn out to be illicit in political campaigning. Please note that because of the range of considerations, final assessments of specific interventions become increasingly complex—too complex, in fact, to present a comprehensive evaluation of any normatively challenging intervention in the confines of this chapter. But I will provide some sketches. My primary aim is to tease out differences between *classes* of interventions. The following is thus not a strike-down argument against direct interventions in every context, but a set of interrelated considerations that place a substantial burden on justifying their nonconsensual imposition on others. A burden heavy enough to warrant a normative dualism as the default rule for the law. Although the argument proceeds from a legal

perspective, the reasons favoring indirect interventions also apply to ethics. Of course, depending on one's further commitments in ethics, these reasons may come to bear differently. But any full ethical theory of mind-interventions has to accommodate them in some way. I thus hope to contribute to the discussion in ethics as well.

Evaluating Interventions and the Ethics of Consciousness

Claims about things of a certain type being preferable to, or less worrisome than, things of another type are comparative. The comparative claim of interest here is that one class of interventions, indirect ones, fare better than another, direct ones.

This of course does not entail that every member of that class is preferable to those of the other. For practical purposes, the main interest does not lie in abstract comparisons between classes of interventions, but in choosing a specific intervention over another under given circumstances. Should parents give methylphenidate to their kids or change schools? Should I take antidepressants or enroll in psychotherapy? Should the state administer drugs or cognitive behavioral therapy to criminal offenders? Should method A or B be deployed (or none)? Answers require knowledge about the pros and cons, costs and benefits, of each intervention. The most important individual element in this are its *effects*, the good and bad, desired and unwanted. To clarify the further discussion, some words about effects upfront. Most importantly, effects are highly specific to individual interventions and therefore do not allow for class-comparisons. There are surely indirect interventions that are much more powerful and with stronger side-effects than direct interventions. Also, effects are of course a contingent empirical matter that cannot be dealt with here. Therefore, the most important aspect in evaluating specific interventions is not in the foreground of the following discussion (but will resurface indirectly at some point). The discussion rather revolves around the question whether there are any *other* normatively relevant aspects, apart from effects and side-effects. To this, the following provides an affirmative answer.

It is nonetheless helpful to understand that assessing effects of a specific intervention is already fraught with difficulties. For one, empirical data is often inconclusive, even with respect to means available for some time (see, e.g., controversies over selective serotonin reuptake inhibitor antidepressants or cannabis).⁵ This might be partially due to shortcomings of industry-sponsored

⁵ For the latest on selective serotonin reuptake inhibitors effectiveness in major depression, see Jakubovski, Varigonda, Freemantle, Taylor, and Bloch (2016).

(pharmaceutical) research. But it is largely a consequence of the challenging nature of the task. Mental effects are often subtle and hard to detect; people respond differently to interventions (e.g., pharmacogenomics), and it proves hard to control for other confounding factors such as cross-effects in studies (they require high *n*-numbers). More broadly, anyone involved in psychological testing is painfully aware of the enormous difficulties in measuring mental states and processes, not only because of their peculiar epistemological accessibility, but also because a comprehensive model of the mind and its parts is missing.

Apart from these general problems of empirical research about minds, further difficulties arise when it comes to evaluating the pros and cons of an intervention. *Kinds* and *strengths* of effects have to be evaluated, and this requires norms. Whether strength is favorable depends, for instance, on whether the intervention is wanted or unwanted by affected persons. Evaluating kinds of effects requires *valuing* mental states, those modified and those newly brought about. It is remarkable that almost hardly any criteria exist for ascribing (dis)value to mental states. Although some appear beneficial and others detrimental, evaluations are regularly not as straightforward as it seems. Even paradigmatic negative mental states as fear or worries are beneficial in some situations. Evaluations require an—as yet, outstanding—comprehensive ethics of consciousness, a theory that provides criteria for valuing mental states.⁶ It faces a range of intriguing questions such as whether positive valenced emotions are ipso facto good, (likely not), whether and to which extent self-critical thoughts are better or worse than self-affirmative ones, or how improvements in one domain may be traded against impairments in another. It may also need to distinguish between mental domains, broadly construed. Different criteria might apply to affective states and to cognitive processes.

Discussions about nosology and mental disorders in psychiatry exhibit some parallels to such an ethics of consciousness, but they derive mainly from a limited set of normative premises around concepts of illness. An ethics of consciousness has to be more comprehensive. It cannot solely rely on subjective appreciations because humans are not well-versed in introspectively grasping how mental elements—thoughts, emotions, moods, dispositions—relate to and affect one another. A science-based ethics of consciousness has to rely on (longitudinal) psychological and phenomenological studies. We are far from having such an ethics of consciousness, its development is a task for neuroethics in the coming decades. Without it, assessing mind-interventions lacks a stable foundation.

⁶ I borrow the term from Metzinger (2010).

Distinctions between Interventions

Misleading Distinctions

Once pros and cons of specific interventions are evaluated, they can be compared to others. One of the key problems is that many interventions are on a first glance incommensurable, they simply produce different effects in different domains. How to compare the mental changes caused, say, by placing children in a different social environment or administering them pharmaceuticals? Common standards need to be created, but they inevitably neglect many facets of the richness of changes caused by interventions.

Instead of such complex comparisons, formal attributes or secondary properties of interventions such as their direct/indirect, natural/artificial, invasive/non-invasive nature are often alluded to. Such formal properties then stand in as proxies for the truly important but hard to assess criteria; they are used as a heuristic as an *attribute substitution* (Kahneman & Frederick, 2002). This is problematic since attributes are often suggestive or misleading. For instance, some interventions are viewed with suspicion because they allegedly cause “permanent” or “irreversible” changes, or “alter the personality” of persons. In light of the brain’s plasticity (and potentially neurogenesis), such claims are empirically questionable. But even if some physiological effects are more durable than others, or touch upon more central characteristics of a person, does it matter? These attributes are used in an evaluative sense, but their normative premises remain implicit. For all we know, if spelled out fully, they may likely turn out to be unpersuasive. After all, higher education or bonding in intimate relationships, to take two examples, likely (and hopefully) cause long-lasting and hard-to-reverse mental (and neuronal) effects, but this hardly raises moral concerns. Permanence, reversibility, or strength of connections are thus no clearly favorable or unfavorable attributes.

Perhaps the most pervasive misleading attribute that needs to be mentioned concerns the fact that direct interventions *rewire* or *change the brain*.⁷ Worries based on the brain changing nature of interventions (at least tacitly) seem to presuppose that other interventions leave the brain unchanged. Such imagined less worrying interventions allegedly only alter the mind (they work “purely psychologically”). Equally misleading are allusions to “rewiring,” which supposedly means creating or strengthening (or weakening or discarding) connections between neurons or changes in higher-level connections of the

⁷ This argument seems impossible to overcome in public discourse. At the day of writing, the science pages of the *New York Times* express excitement over the fact that hypnosis “is not only in the mind” but changes the brain (Goode, 2016).

connectome (Seung, 2012). The glaring problem with such claims is that they rely on unpersuasive mind–brain dualisms that presuppose that some mental effects are causable without changes in the brain or without rewiring neuronal connections. Although the relation between the mind and brain is still a metaphysical mystery, we have ample reason to assume that every mental change is—depending on your favorite theory—caused, realized, accompanied, or supervening upon changes in the brain.⁸ No mental changes without physical changes. Even innocuous interventions change brains: reading these words triggers a cascade of neuronal processes in the brain of you, dear reader, and every time you remember something, you recall (and reproduce) an alteration of synaptic connections in your brain. Meditation, exercise, or psychotherapy demonstrably change the brain. Eric Kandel (2007) remarks that the deeper aim of his lectures is altering brain cells of his students.⁹ Evaluations of interventions—or distinctions between them—based on the mere fact that some change the brain are nonstarters.¹⁰

Equally unfeasible are distinctions between “invasive” or “noninvasive” interventions. In medicine, *invasive* denotes interventions that enter or invade the body, either by puncturing the surface of the skin or “going within” through body orifices.¹¹ Electric stimulation through tDCS, for instance, counts as noninvasive according to this nomenclature. But since all interventions, even the most “psychological”, change the brain, all “go within” and possess the problematic feature of causing changes within the body. Without further explication, these terms are more suggestive than helpful (Davis & van Koningsbruggen, 2013). The same is true for “internal” or “external” interventions: at some point, every intervention is external (to the body) and at another, it causes internal changes (in brain and mind). Many evaluative claims over particular interventions appear innocently unaware of these difficulties and warrant suspicion. However, despite the shortcomings of these distinctions, there might be other, more relevant ones. So, if two interventions, A and B, cause sufficiently similar effects, is there anything else of normative relevance?

⁸ As an introduction to the mind–brain problem, see Chalmers (2002) and Kim (2011).

⁹ For changes in the brain caused by psychotherapies, see Linden (2006); for meditation see, Davidson et al. (2003).

¹⁰ Moreover, the legal literature sometimes speaks about “morphological” brain changes and neuroscience about “structural” or “functional” ones. These attributes are equally inept to designate relevant differences because they primarily depend on the technologies used to detect them. *Morphological* stems from the days when brains were dissected and examined by eye or under microscopes; *structural* and *functional* refer to neuroimaging techniques. These methods measure diverging properties of the brain in different modes and temporal and spatial resolutions but do not denote qualitatively different physiological effects.

¹¹ Cf. Merriam-Webster’s (n.d.) medical definition of *invasive*: “1. tending to spread; especially tending to invade healthy tissue; 2. involving entry into the living body (as by incision or by insertion

Levy’s Parity Principle

I claim that, apart from effects, the causal pathways or (parts of) the *modus operandi* of interventions are normatively relevant and that in virtue of this, indirect interventions are markedly less worrisome than direct ones. This normative asymmetry between interventions stands in contrast to Levy’s parity principle. In his book *Neuroethics*, he formulates a weak and a strong version and adds several specifications (Levy, 2007). I have to leave the subtleties of his profound account aside here. My target is a simpler and more common version of the principle, which can be summarized as follows: Different means to alter minds should be treated legally (or ethically) on a par because there are, apart from effects, no intrinsic differences between means of ethical or legal relevance. Accordingly, what matters in evaluations of mind-interventions is not their nature (direct or indirect, physiological or psychological) but only their effects. Challenging the parity principle thus requires a twofold argument:

1. Presenting a meaningful distinction between direct and indirect interventions that neither falls prey to metaphysical objections, nor contravenes empirical findings about the workings of the mind.
2. Demonstrating why this difference bears normative significance.

Before pointing to the weaknesses of the parity principle, however, I wish to give it due credit. It is a useful *prima facie* test, insofar as it helps to expose the many just mentioned unpersuasive distinctions abound. A good share of the louder voices in public discourse seems biased against the novel, the neuro, and the nonnatural. As a tool to elucidate such biases, the parity principle is commendable and has become an anchor in neuroethics. It presses for justification whenever interventions are treated differently. Here is a legal example. The law regulates various means to alter minds through different frameworks: some drugs and pharmaceuticals fall under strict and tightly enforced international narcotic control regimes, others under more lenient domestic rules, while neurotools such as tDCS and TMS are considered medical devices and are therefore much easier legally accessible. However, since these are all means to alter minds, potentially effective and with side effects, such a piecemeal regulatory approach is unpersuasive, especially as different normative criteria apply to them (Bublitz, 2016). The parity principle calls for justifications of such different treatment. Therefore, it helps to see the general in the particular. It is not false, but too broad and becomes unsustainable in specific cases. Identifying similarities and differences between interventions depends on the level of analysis. The parity principle remains on the surface; it highlights the former and

Proposed Distinction: Sensory Perception versus Purely Physiological Interventions

Making the case for legally relevant differences requires defining more precisely what qualifies as a direct or an indirect intervention. As everything resembles everything else in some sense and differs in another, many distinctions between interventions might be drawn (for other recent approaches see Focquaert & Schermer, 2015; Dahaner, 2019). I would like to suggest the following distinction:

Interventions are indirect or psychological if affected persons perceive (including nonconsciously). Stimuli through their outward senses—vision, hearing, taste, smell, touch.

Direct interventions are those that reach the brain/mind through other, nonperceptual ways (i.e., purely physical-biological processes), such as magnetic or electrical stimulation of the brain. They comprise, but are not restricted to, interventions traditionally considered invasive, including substances that undergo metabolic processes before they cross the blood-brain barrier (e.g., psychotropic drugs).

Admittedly, this is a rough distinction that immediately prompts a set of intriguing questions. So here are some clarifications: These two modalities might not be exhaustive. It is probably useful to consider *genetic* interventions as a further category *sui generis*. The same might be true for bodily activities of persons, such as deep breathing or exercise, which induce psychological changes. They seem distinct from, at least in light of normative considerations, interventions that affect body and mind from the outside, through the administration of substances, physical forces or “energies”, or external stimuli.¹² I am only concerned with such latter interventions. In addition, it is useful to draw further subdistinctions. For instance, pharmaceuticals or brain stimulation methods may have distinct peculiarities that merit special attention. The direct-indirect distinction does not oppose such further distinctions; in fact, finer ones with respect to indirect interventions will be drawn in the following discussion.

With respect to the perceptual nature of indirect interventions, it should be noted that the nature of perception—as well as the senses—are still subject to

¹² Surely, the effects of exercise might be similar to drugs, see Vina, Sanchis-Gomar, Martinez-Bello, and Gomez-Calbrera (2012). However, as it is an illustrative example, when the authors write that “exercise can be considered as a drug,” they are close to a category mistake. Exercise may have similar effects to some drugs in the brain. But that does not negate the many differences between drugs and exercise. One of them, relevant here, might be that the potential effects of exercise are much more restricted than those of drugs and emerge only from internal bodily processes, whereas potent drugs might override such processes.

many discussions. For instance, it is not even clear how many senses human possess and how they should be individuated (Macpherson, 2011). The present argument is confined to the classic five senses that take up information from the external world (taste, sight, touch, smell, sound); it leaves internal senses aside (Macpherson, 2011). This presupposes that exogenous sense experiences are conceptually and phenomenologically sufficiently distinct and discernible from internally generated ones (bodily sensations such as pain, hunger, fatigue). Furthermore, some implications of this distinction may appear counterintuitive: the afternoon tea counts as a direct intervention. Eating chocolate to lift one’s mood counts as a direct intervention if its effects are due to the pharmacological actions of its ingredients, but as an indirect one if caused by its taste.¹³ Direct interventions thus comprise a broad range of interventions, even mundane ones. People directly intervene into their minds on a daily basis. Such a broad definition may appear misguided and neglectful of the specific worries about neurobiological high-tech interventions. However, this objection seems already caught up in questionable preconceptions. Doing justice to mind-interventions requires acknowledging that there might only be gradual differences between a cup of tea and a neurodevice. The broadness of the definition also indicates that reasonable normative assessments require finer differentiations in virtue of additional considerations. But, despite grey areas, sensory perception is the most plausible candidate for a distinguishing criterion that captures normatively relevant differences between the two classes of interventions.

Here is why: let us step back from philosophical questions and Mosaic pieces of empirical sciences and consider some basic facts about humans. They possess a cognitive machinery, physically located in the brain (possibly spread out to other parts such as the gastrointestinal tract and maybe the rest of the body). The cognitive machinery is partially open to the environment. As part of the organism, it relies on bodily processes and requires oxygen, energy, etc. In addition, it is open to the environment in another distinct way: through the senses. Like other animals, humans perceive the external world through their senses. Senses are the organism’s gateways to the outside, its receptors to take up external information. Sensory perception detects and processes external stimuli. Sensuously perceived stimuli can thus be understood as *informational inputs* into the cognitive machinery (which contain sense data). The cognitive machinery is adapted to process such stimuli through a cascade of psychological mechanisms that decode, filter, and engage with incoming information. A subgroup of informational input consists of symbolic and conceptual or communicative inputs, which form a normatively salient category (“speech” or “expression”).

¹³ For chocolate’s effects, see Scholey and Owen (2013).

Both apertures—the sensory and the bodily—can be used to access and alter the cognitive machinery, and both ways involve biological processes. Vision works through electromagnetic waves, touch and hearing work through pressure, and smell and taste work through chemical reactions. Every intervention is thus physical-biological. However, only our sensory modalities acquire and process information. For instance, magnetic or electric forces stimulating brain cells cannot be described as informational inputs (they do not bear informational content), nor is the organism's reaction to them describable as perception.¹⁴ Therefore, direct interventions are no informational inputs into the cognitive machinery, but rather physical alteration of its neuronal substrate. In this view, common distinctions that pitch “purely psychological” against “physiological” interventions frame the contrast from the wrong end. Rightly put, direct interventions are purely physiological whereas indirect ones are physiological *and* psychological.¹⁵

Functions

Let us look a bit more closely at two further distinctions between direct and indirect interventions. The first concerns their relation to the ordinary functions of the cognitive machinery. Irrespective of the perplexing philosophical problems of perception—and so I hope, without getting entangled in the deep controversies around “ordinary functioning” in psychiatry—I suggest that one may fairly say that the ordinary function of the perceptual system is to acquire and process sense data, information from the external world. At least, this is a fairly uncontroversial claim if one assumes, as I will here, that one of the ordinary functions of the brain is information processing. Then, sensitivity of brain cells and brain

¹⁴ If direct interventions (e.g., electric currents) cause sensory feelings (e.g., itching), the latter would count as an indirect intervention because it is sensually perceived. But this effect is a by-product, not the causally effective part of interest. This main effect is produced a direct intervention. Furthermore, direct interventions might be used to convey information. One may use TMS stimulation (“zappings”) as a code (e.g., three failures of working memory may be the signal that working hours are over). This interaction can now be described as *informational*. But on a closer look, there are two interventions. The zapping is a direct intervention because it does not produce its effect (memory lapses) through perception (nor through informational content). In addition, the zapping is used as an informational exchange. In that case, both interventions have to be evaluated on their own.

¹⁵ A usual objection against this distinction states that it still involves a residual dualism as it speaks about psychological processes as distinct from physiological ones. However, if one entertains a physicalist position that does not allow for independent psychological processes, the burden of argument is to show how physical processes instantiate those processes that psychology and cognitive science speak about. As long as they are not explained away (e.g., eliminativist positions), the described psychological processes exist. This remains true even if they are fully reducible to physical processes.

areas to sensory input is a necessary part of its ordinary functioning. By processing such stimuli, the system simply performs this function. In this view, perceptible stimuli engaging the cognitive system thus neither impair its functioning, nor interfere with its integrity.

Whereas one can say with some confidence that one of the functions of the networks of connected brain cells is signaling or conveying information, one cannot, with the same confidence say that picking up and reacting to magnetic fields or electric currents (TMS, tDCS) emanating from sources outside of the skull is. Responsiveness to such stimuli is *not* necessary for the functioning of the cognitive machinery, but rather a contingent feature due to the particular physical realization of the cognitive machinery in humans. Possibly, the neurobiological substrate of the mind, the brain, could have been realized physically in other ways. Functionally equivalent systems running on a different hardware (e.g., insensitive to magnetism) are easily conceivable. Sensitivity to direct interventions is thus not necessary for ordinary functioning. The effects of direct interventions may sometimes even be conceived as distorting interferences with the integrity of the system. When, for instance, neurons increase or decrease “firing” rates because of electrical stimulation of nearby tissue, ordinary signaling is altered. That is the very point of the intervention. Thereby, ordinary functioning is disrupted, at least insofar as signaling no longer works appropriately. Thus, unlike perceptual stimuli, direct interventions may well be described as functional alterations of the system and sometimes as impairments of ordinary functioning. This is an interesting difference with potential normative ramifications.¹⁶ I concede though that the argument may not work if direct interventions restore ordinary functioning (e.g., as in some coerced medical treatments).

Furthermore, I do not wish to purport that this juxtaposition is purely descriptive. Terms such as *integrity*, *should behave*, or *ordinary functioning* are nonneutral ascriptions. Ascribing functions requires reference to a goal (Krohs & Kroes, 2009; McLaughlin, 2000). Whereas the function of technical artefacts depends on the purposes for which they were designed or for which they are used, nature is free from intentionality, teleology, or predestined purpose. In nature things simply happen. In the absence of goals, ascribing (mal) functions to biological systems is problematic. Yet, ascribing the function of perception to the senses, or information signaling to cells, is unsuspecting, as long as it does *not* imply that utilizing the brain for different functions, as additions or expansions, is not (and by extension, ethically or legally impermissible).

¹⁶ It is surely conceivable that neuroscience develops additional sensory modalities for humans—for example, echolocation or senses for electric or magnetic fields that some animals (fish) possess. As long as these are sufficiently analogous to the native senses, especially in the ways in which sense data are processed and integrated, such neosensory inputs would presumably count as indirect interventions.

this requires a normative argument that is yet to come. To anticipate it: persons should be entitled to choose the functions of their organism. They may voluntarily turn their brain into, for example, a receptor for electromagnetic waves, and they may equally deny and close such pathways into their minds. Interveners who nonetheless avail themselves of these pathways then *misappropriate* the mechanisms of the brain (in the sense of making use of them without permission)—and that constitutes an interference with mental self-determination.

Control

Arguments about functions are concededly controversial. Although the previous section reveals a significant difference between interventions, nothing in the more general argument hinges on the point about functions. Sceptics are invited to disregard further references to functions. My main argument draws on another difference between perceptual and psychological interventions: the kind and amount of *control* that affected persons have over it. Control is not a purely descriptive concept either, it designates a relation between a controlling subject and a controlled object. Normatively interesting concepts of control are accordingly tied to notions of a subject, a particularly thorny issue with respect to the mind, which cannot be unpacked here.¹⁷ The term *control* is used variously in psychology (e.g., Hassin, 2005). Nonetheless, even without a clear-cut definition, it is possible to approximate relevant features, distinguish between kinds and degrees of control, and eventually formulate a tentative definition of control that suffices for our purposes. In general, the notion of control over one's mind is broad and includes, for instance, powers and capacities for mental actions. They are often limited. For instance, we often fail to direct our thoughts, or stop the wandering of our minds, and may appear more as passive observers to our stream of consciousness than its directors. A bit provocatively, Thomas Metzinger writes that "for two thirds of their conscious lifetime, human beings do not possess mental autonomy" (Metzinger 2013, p. 14).

¹⁷ It invites questions as to who the subject that controls unconscious mechanisms is. It cannot be the "conscious I." Answers require no less than a construal of the notion of the subject in light of recent findings of cognitive sciences (see Wegner, 2005). While philosophically intriguing the absence of such a concept does not call the suggested normative distinction into question. For our purposes, a concept of the subject that comprises the organism including higher-level mental functions and consciousness suffices. By contrast, accounts that confine control to conscious awareness become implausible in light of the fact that most psychological activities are unconscious. While cognitive sciences may not leave much room for strong metaphysical concepts of subjects, they do *not* call into question differentiations between various kinds and degrees of control.

Our present interest, however, lies more narrowly in control over interventions. My main claim is that the relevant forms of control concern the abilities to detect, engage with, and counteract interventions, and that they typically differ in quality and quantity with respect to direct and indirect interventions. Importantly, the degree of available control negatively correlates with the strength of an intervention. Some interventions are weak and easily resistible; others might be literally irresistible. But strength, as previously mentioned, depends on contingent empirical features of specific interventions and is not a criterion to differentiate *classes* of interventions. The relevant claim in this respect is that the degree of control over interventions *also* correlates with respect to their causal pathways.

Control over Indirect Interventions

Let us first look at indirect interventions, sensuously perceived stimuli. Apart from vision, we cannot shut down our senses or block uptake of information. Through our senses, we are constantly connected to the world (online) to some degree. However, we have some mental control further down the pathway.

A first distinction that suggests itself runs between those stimuli of which persons become consciously aware at some point and those that are merely nonconsciously processed. We have the greatest degree of control over stimuli that rise to conscious awareness. Despite limits, we have mental capacities to engage with these stimuli by giving attention to or distracting us away from them. We can relate to their content, evaluate the information, compare it to pre-existing knowledge and experiences, and categorize or discard it. We have less control over stimuli that remain non-conscious.

This folk psychological approximation is supported by a rough view at the architecture of the mind: although still subject to debate, a picture stable enough for our purposes has emerged. It views the mind as a complex network of specialized modules with different properties that process different types of information in different ways. Some parts of this network run through consciousness; most parts work nonconsciously. It is likely one of the functions of consciousness to enable distribution and integration of information from various modules throughout the network. Through becoming conscious, information is made available to other modules (e.g., global workspace model, Baars, 1997). If this view is correct, one can say that persons have the highest (although still limited) degree of control over consciously available information, because more modules can access and process and engage with it. Nonetheless, all incoming perceptual stimuli are first processed by non-conscious mechanisms. Non-conscious processing is the default mode of operation. The larger share of the myriad of stimuli constantly entering our senses remains nonconscious, only salient stimuli are elevated to conscious awareness. But of course nonconscious information is also further relayed, processed, and engaged with, it may even translate into action

(Bargh & Williams, 2006).¹⁸ All conscious mental activity is grounded in many nonconscious processes “under the hood” that prepare and process information. My suggestion is that even this form of information processing provides some sort of control. Not conscious control, but engagement with stimuli through non-conscious mechanisms and modules. Stimuli are, for instance, checked against available knowledge, predictions, and expectations and are filtered for relevance. Perception itself is likely shaped by thoughts, beliefs, and desires.¹⁹ Persons thus relate to the informational content of such stimuli at least in some way. Because of this, nonconscious stimuli do not prompt random or irrational effects, but rather trigger preset responses, which may often be helpful and adequate. These mechanisms are, to some extent, adaptive and modifiable through learning and novel experiences. Accordingly, I suggest that even these nonconscious processes provide a basal form of control over stimuli, although less control than over conscious stimuli.

There is another angle from which one may approximate relevant aspects of mental control and that complements the conscious/nonconscious distinction. A large body of work from several disciplines converges on dual-process models of the mind.²⁰ Although varying in detail, the common idea is that the mind comprises of two systems that can roughly be discerned: System 1, the evolutionary older one, works fast and parallel, nonconscious, automatic and relies on simple information processing strategies such as heuristics. System 2 works slowly and has only limited capacities but is more deliberate and can access explicit memory. System 1 provides quick and preset responses whereas System 2 enables higher-functioning, reflective thought, and more complex and varied response. Incoming data are first processed in System 1, and some is relayed to System 2. Both systems interact and integrate information; details of their interplay are unclear (Evans and Stanovich, 2013). But despite gaps in the precise understanding of different systems, the dual-process view is illuminating as it explains phenomena such as typical failures of reasoning through properties of various psychological processes. The distinction can also be utilized for present purposes and allows for a tentative conclusion with respect to control: Persons have greater control over stimuli processed by System 2 and less over those processed only by System 1.²¹ How the distinction between the two

¹⁸ The idea of unconscious perception has long troubled philosophers. But empirical science clearly demonstrates its possibility (see Prinz, 2016).

¹⁹ The extent of this influence (“cognitive penetration”) remains hotly contested. See, for example, Stokes (2012).

²⁰ For a good overview see Evans and Frankish (2009) and Kahneman (2011).

²¹ One way of framing the distinction between the system makes this difference in control particularly salient. In analogy to a photo camera, Joshua Greene (2014) speaks of System 1 as the automatic mode and System 2 as the manual mode. The manual mode requires more effort and takes much longer than the automatic mode but allows for more fine-grained control. Although a metaphor, it points in the same direction as the current argument.

systems maps onto the conscious–nonconscious divide is an intricate question and depends on how systems are delineated (Frankish, 2009). Contents only processed by System 1 can rise to conscious awareness, so System 2 processing is not necessary for consciousness. System 2 processing probably requires conscious awareness of some of the stimuli. It therefore seems best to view the distinctions between systems and conscious–nonconscious processing as different dimensions of mental control which overlap but are not co-extensive. But details do not need to concern us here. The main suggestion is that a relevant notion of control can be explicated by drawing on features of the presented picture of the mind. Here is a tentative definition: *the extent of control a person possesses over an intervention correlates with the extent and type of her mental capacities and mechanisms available to engage with it.* As a consequence, control over indirect interventions comes in degrees, with more control over conscious than over nonconscious stimuli and more control over System 2 is activated rather than merely System 1.

Further empirical findings could be adduced to render this picture more precise, but its present form suffices to tease out differences to control over direct interventions.

Control over Direct Interventions

Let us contrast this with control over direct, physiological interventions. Their causal pathways differ, as they do not run through the just described psychological systems, processes, filters, or mechanisms at all; their route is one of biophysical, (chemical, electrical) processes alone. Whether direct interventions are effective solely depends on biophysical events in the brain.²² Roughly, one can put one contrast in this way: direct interventions are mainly about biophysical interactions of the person with the intervention, whereas indirect interventions are about the interplay with the physically embedded information.

Because of different pathways, the kind of control over direct and indirect interventions differs. People can counteract onset and impact of direct interventions to varying degrees. We are accustomed to working against hunger, fatigue, or alcohol, but our powers are limited. The same is likely true for our control over the impact of magnetic fields, electric currents, or pharmaceuticals on the cognitive system. Although not necessarily irresistible, the amount of

²² To be more precise, the received wisdom of drug experiments and psychiatry suggests that effects of psychoactive substances are influenced by “set and setting” (Hartogsohn 2016; Zinberg 1984)—that is, the internal psychological conditions (expectation, motivation) of consumers and context of consumption (perhaps similar to placebo effects). Effectiveness might not only be about biophysical processes alone, it is also about the mental state of the recipient, and in this sense, she may have some residual control.

control over such forces is quite low. Moreover, the kind of control over direct interventions categorically differs from control over perceptive stimuli; the former requires exercising different mental strategies (and, in a sense, mobilizing different mental “energies”). Accordingly, direct interventions often seem to bypass mental control of targeted persons.²³

This aspect seems to be one of the reasons for the personality transforming potential of some direct interventions. Although sensuously perceived stimuli can cause deep effects (e.g., witnessing a traumatizing event), their effectiveness seems to depend on predispositions and the pre-existing state of perceivers to a much greater degree. Sensory perceptions some how have to find traction with person, have to resonate with her, play to her beliefs, hopes, fears, or traits. Accordingly, even drastic indirectly induced personality changes seem to be more in line with the personality and might often be more fittingly described as a reaccentuation rather than a thorough transformation. By contrast, direct interventions are more powerful because they bypass the processes with which indirect stimuli resonate; they directly modify a neuronal configuration and bypass the personality of recipients; more concretely, all the innumerable mental mechanisms that make up who we are. This is illustrated by the fascinating reports of severely depressed patients unresponsive to therapies whose moods are apparently alterable within minutes through DBS (Mayberg et al., 2005; Schlaepfer & Lieb, 2005; for more recent developments see Dougherty, 2015; Morishita, 2014). Conceptualizing such cases by merely saying that DBS is more effective than other therapies is too simple. DBS is more effective precisely because it operates on a different causal route, because it does not resonate with the personality of the patient and her cognitive machinery in a “depressive mode,” but rather reconfigures the neuronal parameters of the cognitive machinery itself.

In conclusion, a crucial difference between direct and indirect interventions lies in their causal pathways. Indirect interventions pass through the ordinary perceptive mechanisms of the cognitive system, direct interventions alter it. Control over one’s mind is limited and comes in degrees. The form of control persons can exert over sensuously perceived stimuli, especially those rising to conscious awareness, differs in kind and extent from control over purely physiological interventions. Control over sensory stimuli can be further

²³ Concededly, this is a rough and oversimplified picture that fails, for instance, to differentiate between sensory modalities. Think about the lack of control over sensory perceptions of smells, due to the fact that olfactory stimuli are not relayed to the thalamus. Nonetheless, if one agrees with the idea that we possess less control over smells than over other perceptible stimuli, one affirms my general suggestion that control depends on pathways of stimuli. Moreover, both direct and indirect interventions can trigger further equally hard to control effects (e.g., smells evoking memories). But such further effects are not our concern here.

differentiated in virtue of conscious/nonconscious and System 1/System 2 processing.

Of course, the overall amount of control of a person over interventions depends on additional factors including her general abilities, further circumstantial factors, as well as the strength of interventions. While these empirical variables become relevant in evaluations of specific methods, they do not allow identifying principled differences between classes of interventions and must be held fixed for our present inquiry. Thus, *ceteris paribus*, persons have more control over indirect than direct interventions.

Normative Relevance of Differences

Against this backdrop, we can turn to the normative part: Why should these differences matter? Answers require normative criteria. A legal, rights-based perspective focuses on legal relations between persons, more concretely, on legal relations that pertain to minds. What is the legal relation of a person to the mind of another; which actions potentially altering another’s mind are (im) permissible? The legal relation of a person to the mind of another is secondary to a more fundamental relation: the one between a person and her own mind. Are persons legally entitled to alter their own minds, can others—or the state—keep them from doing so? And what rights do persons have against unwanted alterations of their mind?

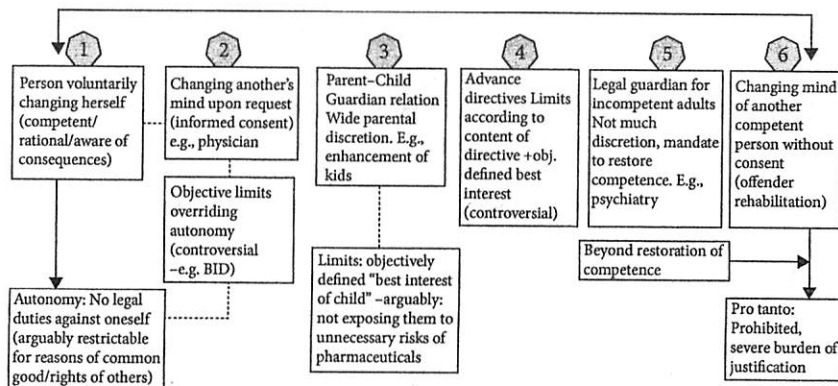
The law rarely addresses these intriguing questions in abstract. While many legal doctrines regulate how to treat bodies, our own or those of others, few explicitly concern minds.²⁴ Even regulations of mind-altering tools such as drugs are not formulated in light of a general theory over rights to minds, but are piecemeal legislation mainly based on considerations of harmfulness.²⁵ The mind is a largely uncharted area of the law. Filling the void and evaluating mind-interventions adequately requires developing a full-fledged legal theory over rights to minds. In the following, I shall present some cornerstones, especially a (human) right to mental self-determination, which is not yet accepted in many jurisdictions, but should be incorporated in the catalogues of international human rights. This is the normative premise of the argument.

²⁴ One might think that rights to the body encompass rights to minds, because minds are realized in brains. However, legally, rights to the body do not fully capture interventions into minds. For several reasons, distinct doctrines for the protection of the mind have to be developed (cf. Bublitz & Merkel 2014).

²⁵ The international drug convention, for instance, do not even consider a right of users to alter their minds (Bublitz, 2016).

Different Legal Relations

Before elaborating on the substance of the right, I wish to make a methodological suggestion: assessments of interventions should pay attention to different formal relations between interveners and targets (cf. Merkel, 2007). Because of them, generalizations, and analogies might not be easily drawn. The failure to observe these differences befalls many contemporary discussions of the issue. Senders and receivers of mind-altering interventions can stand in different formal relations to another, and these relations affect corresponding rights and duties. Next is an overview of the most important types.



In the first case (upper left), a competent person voluntarily alters her own mind. Here, intervener and receiver are identical, whereas all other cases involve (at least) two persons. Accordingly, the first case concerns legal limits to alter one's *own* mind, whereas the other cases concern rights and limits to alter *other* minds. In a legal view, these cases differ categorically because a right barring mind-interventions only obliges *other* persons, not rightholders themselves.²⁶ Consequently, the only question in the first case is whether there are any limits to mental freedom (and how they can be justified). In cases concerning *other* minds, the specific relation between intervener and recipient strongly influences whether and to which extent interventions are permissible. Different substantive rules and considerations apply. They derive from general considerations, are not peculiar to mind-interventions and are controversial in all case types. The reason for introducing these distinctions is of methodological nature: Because diverging rules apply to different relations, the dissimilarities between direct and indirect interventions may play out differently in various case types. The most salient example concerns the paradigmatic case of enhancement: A competent person voluntarily enhancing herself (1) is in

²⁶ One of the reasons lies in the nature of rights: rights and duties against oneself arguably cannot exist (as rightholder and dutybearer would be identical).

light of mental self-determination less unsettling than parents enhancing children (3), which is less unsettling than criminal justice systems "enhancing" competent offenders against their will (6). Worries over direct interventions in the latter case may not apply to the former. And reversely, even though one comes to conclusion that direct interventions are permissible in voluntary self-enhancements (1), it does not follow that they are equally so in (6). Case-based reasoning in neuroethics often fails to distinguish between various case types. It could gain precision by becoming sensitive to these distinctions. The most problematic cases in which differences between interventions are mainly relevant are unwanted interventions (especially 5 and 6). I will only be concerned with interventions into other minds (i.e., not case types 1, 2, and 4).

Rights against Mind-Interventions

We can now turn to the substance of affected rights. The key one, I suggest, is the right to mental self-determination or, as often called in the United States, cognitive liberty. Although not (yet) acknowledged by most legal systems nor found in legal textbooks, I claim it is a fundamental human right. It draws upon the guarantee of freedom of thought, enshrined in many human rights treaties²⁷ and is a logical prerequisite to many more specific human or constitutional rights.²⁸ It may be viewed as a logical expansion of the right to bodily autonomy and flows from the idea that individuals should decide self-regarding matters for themselves, a basic premise of liberal legal orders (widely endorsed in western jurisdictions). It would be hardly conceivable if autonomy stopped at what is most central to the person: her mind. Mental self-determination has to be the starting point for legal assessments of mind-interventions. But as here is not the place to provide a deeper theoretical justification for the right, I have to direct the reader to discussions elsewhere.²⁹ Let me, however, sketch my view on content and scope of the right: as the name suggests, it confers upon its holder the liberty to self-determine or control what is in and on her mind. This entails the permission to alter one's own mind at will. It further implies, by the logic of norms *that others do not have claims over the content of the rightholder's mind*. No one can legally demand that the rightholder entertains a specific thought, feels a certain way, or possesses or exercises particular mental capacities. The importance of this claim will become evident in a minute. Additionally, the right implies a claim against others to refrain from altering minds of rightholders against their will (therefore, interventions in case types 5 and 6 are pro tanto impermissible).

²⁷ E.g. Art. 18 Universal Declaration of Human Rights; Art. 19 International Covenant on Civil and Political Rights; Art. 9 European Convention on Human Rights.

²⁸ Boire, 2001.

²⁹ See Bublitz/Merkel, 2014; Bublitz (2014a,b).

Three clarifications are warranted. The right may have limits. Moreover, rightsholders have the power to consent to mind alterations (relevant for case types 2 and 4). Furthermore, a right to mental self-determination does not imply or presuppose that humans are ideally self-determined or self-controlled beings. We are not. As a legal right, mental self-determination pertains to relations *between* persons, not to psychic or bodily forces within a person. It guarantees negative liberties against interferences by others.

To render the scope more precise, let us image how the world and social interactions would become if the duty to refrain from altering other minds were strictly observed. Then, we would be obliged to refrain from sending mind-altering stimuli to others, which means that we could not even speak to one another without obtaining prior permission. Such a strict construal of the right is implausible. The law is, after all, interested in finding norms for a reasonable social order of which cooperation and communication are essential elements. Blanket prohibitions to alter other minds are thus neither practically conceivable nor politically desirable. While withdrawal from social interactions might be a legitimate conception of a good life, and social interactions should not be imposed on those who seek solitude, a Thoreauen lifestyle cannot serve as a model for normative orders for contemporary societies. Everyone who seeks to transcend solipsism desires and depends on social interaction. The law's task is to enable and promote meaningful and useful forms of interaction while curbing those forms that undermine mental self-determination in a troubling way. Therefore, the scope of the right against mind-interventions has to be confined to interferences that undermine mental self-determination to a degree that fails a test of what is reasonable in a highly cooperative, interactive, and communicative society. And this, I claim, is true of interventions that bypass mental control. Most of the myriad of stimuli entering our senses everyday, by contrast, fall outside of the scope of the right. In light of this normative premise, the relevance of mental control and integrity—and the direct versus indirect distinction—becomes evident. The more control, the better. In addition, mental self-determination entitles persons to define the functions of their cognitive machinery for themselves. Those wishing to utilize their brains as receptors for magnetic fields outside of their skull may ascribe this function to them. Magnetic receptivity then becomes a feature of their cognitive system. Those, by contrast, who do not wish to open such a pathway into their minds may reject this function. Using this pathway against the will of the person amounts to illicitly appropriating another's cognitive system.³⁰

The upshot is this: everyone has a right against unwanted mind-interventions. It correlates with a duty of interveners to respect mental control and to refrain from altering minds of rightsholders. Interventions undermining or bypassing mental control, or appropriating functions of other brains, are particularly troublesome.

³⁰ The same cannot be said with respect to ordinary sense functions. If person A talks to B against the latter's will, it seems implausible to say that A has misappropriated B's sense functions.

Let us tie this together with the previous analysis and recapitulate: Interventions can be differentiated in terms of control. Persons have more control over stimuli processed by System 2 that rise to conscious awareness, less over those processed nonconsciously or by System 1 only, and even less control over direct interventions. In light of the right to mental self-determination, the latter are thus more problematic than the former. Respecting mental self-determination thus commands refraining from altering minds through direct interventions or through problematic control-undermining indirect interventions (only nonconscious or System 1 processing).

Rights to Intervene into Other Minds

At this stage, I hope to have shown why—in light of a reasonable construal of the applicable norm—the law should draw distinctions between interventions. Yet so far, only rights *against* mind-interventions have been taken into account. Perhaps, however, interveners may invoke rights *favoring* interventions. After all, mind-interventions are social interactions. Rights against interventions are only one side of the equation. If a conflict between rights for and against interventions emerges, rights have to be reconciled through the applicable method (e.g., balancing). This conflict plays hardly any role in current debates, but only its appreciation allows grasping the full scale of the normative issues at stake. Which rights could interveners invoke? Potentially several, especially in special circumstances (e.g., case types 3–6). But rather than elaborating on such specific rights, I wish to draw attention to another structural difference between direct and indirect interventions in this regard. The following discussion might get a bit tricky. In the previous section, I argued that interveners do not possess rights over minds of targets (“No claims over rightsholders’ minds”). Therefore, interveners cannot *justify* mind-altering interventions by a claim that the target is under a legal duty to be in the mental state which the intervention produces. However, interventions might be justified by *other* considerations, primarily by rights permitting interveners potentially mind-altering *actions*. Free speech is the prime example. It can be conceived as a permission to send potentially mind-altering stimuli of a particular form, communicative stimuli. While it entitles interveners to send stimuli, it does not guarantee that the message be heard or appreciated. Senders may speak, but recipients do not have to listen. The key to solve the conflict between free speech as a permission to send stimuli and mental self-determination as a right against receiving them lies in distinguishing between *actions* and *effects*. Intervenors have a right to speak but no claim that speech produces the desired effects in listeners. Free speech may justify communicative mind-interventions if they turn out successful, but it does not guarantee the conditions necessary for success.

More has to be said about the scope and strength of free speech and finer distinctions be drawn with respect to particular situations (e.g., “captured audiences” who cannot escape exposure to unwanted speech), means and

purposes of communication (e.g., advertising), and atypical cases. It is worth noting here that usually only indirect interventions give rise to such conflicts since direct interventions are usually *not* protected by rights such as free speech because interveners do not pursue (legally) legitimate ends with their action. They primarily seek to alter targeted minds—to which they are *not* entitled because of recipients' right to mental self-determination. To illustrate: regularly, the only aim and interest of an intervener spraying a psychotropic substance such as oxytocin in a room or applying a magnetic field to another's forehead is altering the mental world of affected persons. Targets can reject tampering with their minds because of their right to mental self-determination. They could not, by contrast, bar interveners from speaking to them (because of free speech of interveners). More generally: The action of directly intervening (e.g., pushing a button) is, by itself, regularly not valuable for intrinsic or social reasons; whereas the action of indirect interventions (expression) is. Accordingly, a structural difference in the justification of direct and indirect interventions emerges: some indirect interventions are exercises of strong rights; direct interventions are usually not.³¹

In cases in which a right to the intervening actions such as free speech exists, it has to be balanced against the countervailing right to mental self-determination. This requires context-specific considerations. Effects and strength of interventions play a key role, weaker ones undermining mental self-determination to a lesser degree are easier justifiable than stronger ones. But strength is not the only decisive factor. Highly effective, even irresistible interventions which are doubtlessly permissible are easily conceivable. The prime example is a persuasive argument that cannot but make (a rational) listener change her opinions or beliefs. This is the "forceless force" of the better argument, and it differs substantially from the force of, for example, a chemical intervention targeting opinions or beliefs. Rational and justifiable beliefs require a particular mode of acquisition and transformation: reason or evidence (Crutchfield, 2016). Of course, communication regularly involves more than rational exchange of reasons. The law draws distinctions among various forms (e.g., manipulation, coercive persuasion, undue influence). Communicative interventions fall on a spectrum. Habermasian ideal speech situations in which statements are formulated free from emotional influences or distracting associations in contexts devoid of power lie on one end. There, only better arguments prevail. On the opposing end lie communicative forms deploying psychological trickery and exploiting psychological weaknesses of recipients. Lines of moral and legal significance run even between these poles. A parity principle obfuscates them.

³¹ Note that in exceptional cases, the norm "No claims over minds of another" is false, because other persons do have such claims, most notably, parents over children (case type 3). Parents not only have a right; they are even under a *duty* to shape their kids' minds (e.g., education and development of mental skills). The same might be true for the state or guardians with respect to noncompetent persons, or even, most controversially, with respect to criminal offenders. These case types require finer discussion. Here, I only wish to reiterate the previously made methodological point that such atypical cases do not allow for generalizations.

Where precisely the boundaries between permissible speech and undue persuasion are best drawn is a challenging question that dates back to Greek debates over the ethics of rhetoric. Instead of providing answers, I can only note that it is a shortcoming of legal theory and philosophy to not offer more concrete guidance about "undue influence" or "illicit manipulation," concededly a complex matter in which empirical and normative considerations are deeply interwoven (see Coons and Weber, 2014). Especially (social) psychology, cognitive sciences, and "marketing studies" have produced a wealth of findings that needs to be analyzed in light of ideas of mental self-determination and free speech, such as the role of emotions in attitude formation and communicative strategies appealing to them (see Cialdini, 2007, 2016). History has shown the effectiveness of propaganda by indirect means stirring emotions of the masses, and instilling fear and insecurity still seems to be a winning strategy of political campaigns at this very moment in the United States and Europe.³² In the dawning age of "posttruth" and "alternative facts," the primacy of opinion change through rational discourse tends to be replaced by vindication of intuitions, ideologies, and sentiments. We should insist on the importance of justifiable beliefs and the primacy of argument in opinion formation as it forms the bedrock of societies, which wish to engage with each other on, by and large, rational ways. Therefore, a default policy for a social order of reasonable citizens has to hold that beliefs should be altered through evidence and argument and not by interventions bypassing reasons and reasoning. Challenging work about the normative limits of communicative influence awaits to be done, and some sophisticated forms of persuasion (e.g., in advertising or political campaigning) might have to be restricted more strongly. But that opens a long debate about the value and limits of free speech, which would lead us astray here. But it indirectly proves my point. To reiterate: a key difference between direct and indirect interventions is that restricting the latter leads into thorny issues of free speech; restricting the former does not.

Free speech is the most salient example of the asymmetrical relation between direct and indirect interventions and suffices to cast doubts on the parity principle. Taking parity seriously entails losing criteria to distinguish between communicative and other interventions. But what about other forms of influences via the senses that we send and receive all the time? Given their ubiquity and manifoldness, it is impossible to survey them here. But the structure of their assessment is the same: rights of interveners have to be reconciled with mental self-determination of recipients. A brief and basic example: Imagine an ordinary conversation between A and B in which the former is strongly influenced by the latter's appearance—her style, dress, look, posture, tone of voice, and social standing. While A remains largely unaware of these influences, B uses her charms intentionally. These nonrational

³² See Bargh (2017) for some astonishingly simple experiments that change political attitudes by evoking emotions of fear and security.

factors change A's mindset, so that he finds her arguments persuasive. B's noncommunicative influences are no instances of free speech. Do they violate A's mental self-determination? Arguably not. The way a person appears and presents herself to others and behaves are expressions of her personality and protected by rights to the person. They may justify B's impact on A's mind.

In this way, other noncommunicative influences need to be analyzed. Further rights may come into play, as well as further consideration such as implicit consent. In many cases, a right to send influential indirect stimuli might be absent, and in some instances, they clearly run afoul of mental self-determination (e.g., subliminal stimuli, see Dijksterhuis, Aarts, and Smith, 2005). Many of them, however, are also inevitable manifestations of social life. They unavoidably accompany cherished forms of social practices and thus cannot be restricted without curbing these practices, which seems unattractive in many cases. Every mundane communicative intervention is befallen with nonrational influences just mentioned. But, nonetheless, whereas indirect interventions regularly accompany desirable forms of social interventions, direct interventions seem to be only rarely inevitable by-products of desirable social interactions.

Consequences

What follows from this? In light of mental self-determination, indirect interventions fare better than direct ones for several interrelated reasons: Indirect interventions leave the integrity of the cognitive system intact. Regularly, they neither misappropriate functions, nor bypass control capacities of affected persons. Furthermore, insofar as they are instances of communication, they enjoy free speech protection. Direct interventions, by contrast, bypass control capacities, may misappropriate functions, and legal rights justifying them are regularly absent since no one has claims over another's mind. Indirect interventions of noncommunicative nature require further context specific evaluation. As long as they inevitably accompany ordinary interactions, they are likely the kind of mutual influence that social life entails and that escapes reasonable regulation. Still, some forms might be illicit. This allows ranking interventions in an ascending order of normative concern, with prior ones being normatively preferable to subsequent ones. The following are some paradigmatic instances.

1. Perceptual stimuli, available for conscious and System 2 processing, of communicative nature
2. Perceptual stimuli, available for conscious and System 2 processing, of noncommunicative nature but protected by rights
3. Perceptual stimuli, only available for nonconscious or System 1 processing, inevitably accompanying desirable social interactions (e.g., priming)

4. Perceptual stimuli, only available for nonconscious and System 1 processing, avoidable (e.g., subliminal stimuli)
5. Direct interventions, avoidable

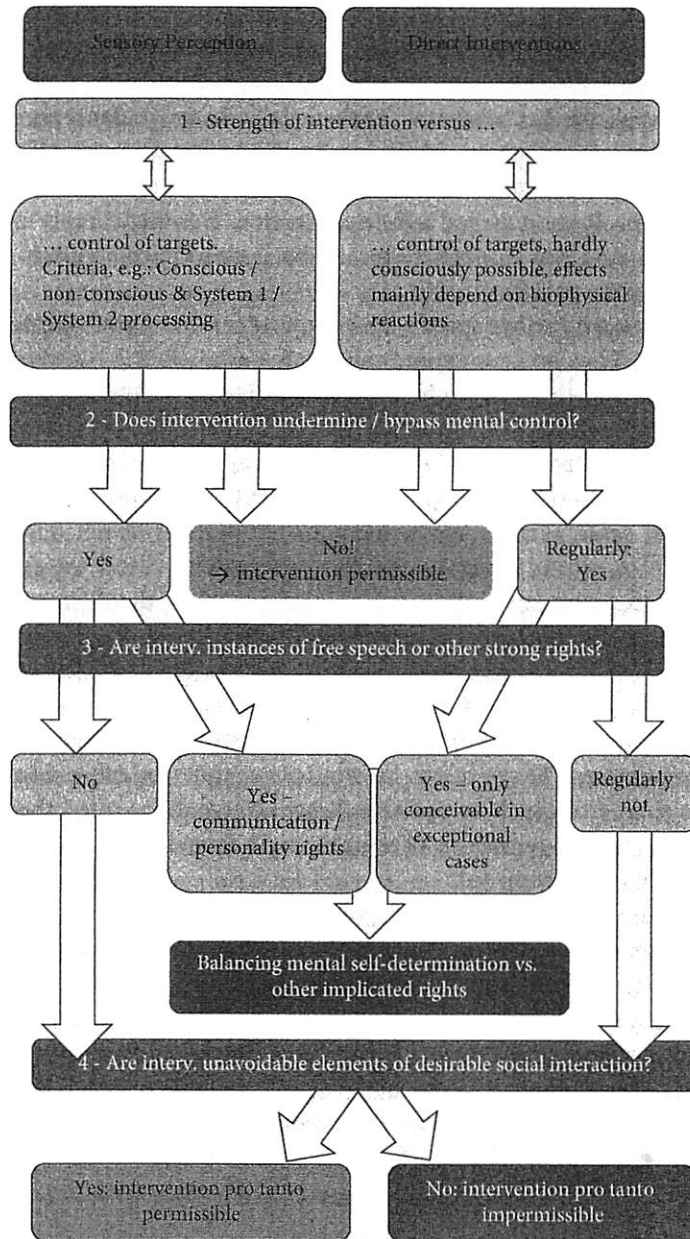
Herewith, some progress to the overall aim of developing a taxonomy of mind-interventions is made. Of course, empirical facts such as strengths of interventions and further normative criteria have to be added to complete such a taxonomy. But it allows for some conclusion. First, suppose a prospective intervener (e.g., policymaker, advertiser, or therapist) asks for guidance on whether and by which means they should change the mind of someone else? The parity principle denies differences. But that surely does not reflect the best answer. The advice must be, Respect mental self-determination. This might imply to not intervene at all. And if so, through the means that leave recipients with the highest degree of mental control.

Second, we can further specify the prospects for direct interventions into other minds. They are, *cum grano salis*, only permissible in a narrow range of cases: For one, there might be cases in which interveners do have a right to alter minds (either to the action or action and effects), because the right to mental self-determination is *limited*. Whether and where limits may run is a question I cannot address here. But, if at all, limits will be restricted to exceptional cases, such as legally incompetent persons or rehabilitation of offenders.³³ A second category concerns unavoidable direct interventions. For instance, eating at a restaurant regularly involves direct interventions (through psychotropic properties of food—think only of glucose; Wenk, 2015). While restaurants should surely respect mental self-determination of customers (and inform them about possible psychotropic effects of meals), this is a good example of socially accepted (and low-intensity) direct interventions. There might be similar examples in other domains of life. But apart from such exceptions, direct interventions into other minds are impermissible—regardless of their strength (provided they are effective above a *de minimis* threshold).

Where finer lines should between indirect interventions should be drawn, especially between sophisticated forms of communicative persuasion involving appeals to emotion, depends on further value decisions. The key question is this: What should be the default mode of engaging with other minds, and which modes undermine mental control to a degree so disquieting that the legal system should step in? After all, as a guiding line, it suggests itself to bar those interventions which severely undermine mental self-determination and

³³ I argue against such coercive interventions in the minds of offenders in Bublitz 2018. For opposing views, see Douglas (2014, 2018).

fail to serve legitimate ends. This entails banning direct interventions. And here we have arrived at the asymmetry between interventions that this chapter sought to demonstrate. The following figure shows the relevant questions schematically:



More Complex Interventions

The foregoing analysis was confined to *single* interventions. More complex interventions may comprise a range of various stimuli. For instance, therapies in psychiatry often involve a combination of drugs and talk therapies. Or, more sinister, consider gaslighting, or brainwashing, a vague concept for a set of severe manipulative interventions exploiting mental weaknesses and instabilities (Taylor, 2006). Although not necessarily involving direct interventions, the latter are paradigmatic examples for illicit attacks on mental self-determination of victims. Similarly, methods such as aversive conditioning based on uncontrollable learning mechanisms that associate unrelated stimuli (e.g., aversive conditioning in *Clockwork Orange*) may infringe upon mental self-determination. More benign, psychotherapies may trigger hard to control emotional dynamics or nonconscious processes such as (counter)transference (Lemma, 2016).³⁴ Likewise, sophisticated audiovisual installations or immersive virtual realities can be powerful influences. Conversely, experiments showed that sensory deprivation of external stimuli had massive effects on the mind (Grassian & Friedmann, 1986). Also, indirect stimuli might be harder to control under specific external (social) circumstances that weaken mental control, from time pressure to situational factors. In combination, interventions consisting of several, by themselves, innocuous stimuli can amount to massive intervention into minds, and, if imposed involuntarily, mental self-determination speaks in favor of their ban.

Environmental Interventions

Against this backdrop, we can assess some recently discussed cases. The first concerns advertisements in public places. Imagine a billboard featuring the actress Jennifer Aniston. Perceiving it creates a visual representation of Aniston in the perceiver, which triggers further thoughts about her. Moreover, Nicole Vincent suggested a garden in which the neuropeptide oxytocin is released, making visitors more open, trusting, and more intimately conversing. Thomas Douglas challenges my distinction with a case in which the walls in an institution (a ward or prison) are painted in a color that calms and soothes people, who remain unaware of the influence (Douglas, 2018). As the paint manipulates persons to the same degree than direct interventions may, Douglas considers the

³⁴ Transference is a process in the relation between patient and therapist in psychoanalytic therapy. In theory, the patient transfers feelings and attitudes of older interpersonal relation on the therapists. Therapies may make use of this nonconscious process.

direct–indirect distinction unpersuasive. Finally, a common objection against the distinction refers to nudging through choice architecture.

Let us assume the interventions in these cases are effective. They are peculiar because they are not confined to one-to-one situations but alter public places and potentially affect a broader audience. Let us call them *environmental interventions*. Normatively, they raise questions as to how environments should be designed, by whom, and about the extent to which affected persons may legitimately complain about the world being designed in a specific way. That environmental conditions alter behavior is among the central tenets of behaviorism. To Skinner and in a somewhat strange relation, the leftist movement of the 1960s and 1970s, the environment was the decisive causal factor (Skinner, 2002, cf. Wheeler, 1973). Interventions accordingly aim at changing the environment (including social conditions), which may then change people. Even though the scientific demise of behaviorism was mainly due to its overemphasis on the environment, it is beyond doubt that environmental stimuli can effectively alter behavior (Chomsky, 1971). And, on my account, normatively, evaluating environmental interventions requires balancing countervailing rights and interests in the manner previously outlined, with the difference that social or public considerations come to the fore.

Take the Aniston billboard. Perceiving means picking up information from the external world, to which we are constantly connected. Perception thus necessarily implies that minds of perceivers may be altered through perceived stimuli. Thus, our existential mode of being connected to the world through our senses brings alterations of our minds with it. Normatively, this suggests (although not necessarily entails) that perceivers usually cannot complain about alterations of their minds through perception. As long as perceptions are veridical, perceivers acquire knowledge about the world through perception. They may only complain about the existence of particular stimuli (or about the existential mode of being connected through the world via perception, but that is not a legally relevant complaint).

Whether complaints are warranted is a normative question that has to accommodate several aspects, among others, strength of stimuli and effects on mental self-determination. In addition, it involves normative considerations over the design of public spaces. Because people share the external world, no individual can claim priority of her particular interests over those of others, and that includes her mental reactions to perceiving the social sphere. The design of public places is therefore a *res publica*. Options range from a low stimuli environment—say, painting everything in grey—to high-stimuli environments such as Times Square. Communities have good reasons to choose the latter, to paint walls colorfully or put up boards for communicative exchange. In the design decision, mental self-determination of perceivers has to be recognized, but only as one

among several relevant factors. Accordingly, beholders of a billboard cannot complain about having a visual representation (i.e., what perceiving implies), nor about perceiving female faces in public or Aniston in persona. They could only complain about the particular representation on the board. The setback of this intervention into their minds has to be weighed against the legitimate democratic prerogative over designing public spaces, and the interests putting up such billboards pursues. In this case, the effects on mental self-determination through the Aniston stimulus are so trivial that community interests to afford spaces for visual communication prevail. However, not all forms of public advertisements appear beyond scrutiny. More intensive and systematic interventions—imagine a city plastered with Aniston pictures, for instance—may violate mental self-determination.

But, in any case, billboards contrast with direct interventions: In experiments with epilepsy patients, neuroscientists discovered “concept cells” (as it were of Aniston). These findings suggest that concepts are stored in individual cells, which function as building blocks of memory (Quiroga, 2012; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). Imagine that instead of a billboard, a device is installed that stimulates the concept neuron of Aniston whenever a person walks by, triggering the same mental effects as the board. This direct intervention interferes with mental self-determination as there is no reason to have a visual representation of Aniston walking by that spot. Further, installing the device does not serve legitimate interests of the community (here, again, the right to mental self-determination bars legal protection of interests in other persons entertaining specific thoughts). The difference between the billboard and the device—although both have identical effects—is that good reasons of the community may speak for putting up the former, but none for the latter.

Surely, one may wonder why we should not enrich the environment with such direct interventions. This brings us Vincent’s oxytocin garden. There is no principled objection against such places as long as affected persons consent to exposure. Visitors aware of the oxytocin may be deemed to consent by entering. Without consent, oxytocin impinges upon mental self-determination. We can also view the Aniston case through the lens of consent. A person walking with eyes wide open through an urban environment can be deemed to consent to the impressions she perceives, inasmuch as they remain within the range of the expectable. Likewise, if instead of oxytocin, the garden was full of splendid flowers inducing a warm and opening mood, perceivers lack grounds to complain as this is a typical and expectable response to an environment designed in a socially adequate way.

Consent is also one of the problems in Douglas’ case. The wall color is a sensual indirect intervention. In fact, a shade of pink, Baker-Miller pink, is suspected of reducing aggressive behavior. Some prisons cells were painted in the tone, but

its effectiveness is unclear (Genschow, 2015). More generally, physical environments, spaces, architecture, and color can exert influence, in the way that visiting awe-inspiring buildings does. But while effective, I suspect that strength should not be overstated. According to the previously described distinctions, sensory perception leaves perceivers with some form of control (this might be testable empirically). The arising normative question is how places as prisons should be designed. As any design of such institutions will inevitably have some effects, they are, as such, not avoidable. An inhospitable clinical setting would also effect people. So, if one designs a prison which alters minds of inmates in any case, why not in a way creating an atmosphere in which people feel secure, comfortable, and at ease? If all spaces affect people, designing them conducive to their functions seems appropriate. Particularly influential forms are nonetheless troubling. The example in point is the Panopticon as described by Foucault (1995). Cells of inmates are constantly visible because of the setup of the building. The constant exposure to surveillance lets inmates internalize a feeling of being surveilled and, thereby, alters minds and behaviors. However, these effects touch upon further issues such as privacy in corrective settings. To avoid these additional complexities, let us assume Baker-Miller pink has considerable mind-altering effects quite unlike ordinary colors, is effective without awareness, and can neither be avoided nor resisted by affected persons. Then, the paint amounts to a constant indirect intervention and may interfere with mental self-determination. However, exposing prisoners to it might be justifiable by the special conditions of incarceration, and this seems to do some of the argumentative work in the example. While prisons lack claims over minds of inmates, they have a right—and a duty—to enforce compliant *behavior* that includes, in perilous situations, physical restraints or even direct mind-interventions (e.g., a sedative). One might argue the institutional interest in reducing aggressive behavior justifies less-invasive means such as a calming paint. The crucial aspect is thus not the perceptible nature of the intervention but its inescapable omnipresence and potential justifiability.

Finally, consider a basic case of choice architecture (e.g., positioning and pricing of goods in a supermarket). To Sunstein and Thaler, such nudges do not interfere with liberties or rights of affected persons at all (Thaler & Sunstein, 2008). This is why their idea of “libertarian paternalism” is so intriguing. If they were correct, the legal difference to direct interventions is evident: the latter surely interfere with rights such as mental self-determination. However, according to my construal of the right, choice architecture does appear worrisome. Positioning of goods in supermarkets is not a matter of free speech. It may be an expression of shop owners’ rights to property or business that allows them to set up stores how they see fit. However, they are obliged to respect mental self-determination of customers. Choice architecture falls within the tension of these

rights. While one may say that placing goods in a particular order only rearranges existing stimuli, it is also an intentional attempt to steer decisions and behavior in ways of which customers are not aware. Respecting mental self-determination implies not exploiting mental weaknesses. Balancing rights may yield the compromise that customers have to be informed about these influences. Awareness might empower them to consciously (System 2) reconsider their choices (although indications are that people also succumb to nudges they are made aware of; Sunstein, 2016). At the same time, a duty to inform about choice architecture does not severely infringe rights of shop owners. They can still arrange goods as they please. Yet again, choice architecture differs from direct interventions (e.g., releasing an odorless substance with similar mind-altering effects). Directly influencing choices is something to which shop owners are not entitled, because it does not serve legitimate interests in running their business. Their only aim is altering choices, and that, as such, is off limits to them due to customers’ right to mental self-determination. To sum up, while some of these environmental interventions give grounds for concern, they often require further context specific considerations. In any case, they do not disprove relevant differences between direct and indirect interventions.

Reply to Levy

With the previous discussion in place, let me finally address some objections Levy raises in Chapter 2 of this volume against (an earlier version of) my account. I concur with his criticisms of ill-founded distinctions between interventions and with large parts of his present argument. Still, and somewhat surprisingly, we reach diverging normative conclusions. Broadly, Levy seeks to dismiss the relevance of the distinction by showing that many indirect interventions are worrying precisely for the same reason as direct interventions are—namely, because they undermine control or alter central characteristics of persons. In such general terms, I agree. Our disagreement, though, seems partly due to the methodological point I raised earlier. Levy presents the example of a moderately depressed person who exposes herself to sunlight or adds iodine to her diet. I agree that there is nothing to demur to absorbing sunlight or iodine. Voluntarily changing one’s own mind through direct interventions is often unproblematic and the right to mental self-determination not even implicated because it only obliges others (it does not create duties against oneself). However, if persons are exposed to sunlight or mind-altering foods against their will, the right kicks in (imagine a prison inmate is put on a mandatory diet to reduce aggression). Analogies between cases of voluntary self-change to nonvoluntarily interventions are misleading (see the distinction between *case supra*).

Furthermore, in the foregoing chapter, Levy objects to the “lack of control” criterion for two reasons. For one, persons may have “the chance to assess the new mental state” produced by direct interventions in the same ways as those produced by indirect ways. I agree, but we seem to talk past each other. My argument pertains the kind and degree of control persons have over interventions from the moment of first contact until their effects set in (where that endpoint is remains concededly a bit blurry). It does not pertain to more downstream effects and how (easily) reversible they are (largely a matter of strength). Levy further argues that people may not have full control even over consciously perceived stimuli. He notes in the previous chapter that “consciousness of stimuli is not sufficient for control over how we respond” This well-chosen example demonstrates that we can be “conscious of words, but not of the effects the words have” on us. Again, I agree. But this does not undermine the relevance of the direct–indirect distinction. I do not wish to suggest, nor does my account presuppose, that humans are anything near ideally self-controlled agents. Large parts of our psychological operations remain opaque to us, we post hoc rationalize and confabulate. Such weaknesses render us vulnerable to exploitation by others. Yet, the existence of such weaknesses does not imply or suggest the truth of the parity principle. The priming examples nicely illustrates this. Susceptibility to conceptual or perceptual priming is a feature of our psychological system, priming occurs in ordinary situations and conversations. Because it is effective, one may call for an “ethics of priming.” Attempts to regulate priming, however, face the problem of its ubiquity and raise free speech–related difficulties. Some priming effects (e.g., in a rhetorically skilled speech) seem to be legitimate exercises of this right. More generally, unless one is prepared to severely obstruct communication (e.g., by obliging speakers to choose words in ways devoid of priming effects, if possible at all), it largely escapes regulation. Perhaps special forms or contexts might be regulatable (e.g., masked priming, advertisement), but priming as such is just not restrictable across the board. So, whereas Levy is right that priming is not beyond concern, his conclusion is questionable: the existence of hard-to-regulate control undermining influences such as priming does not imply that other, equally effective, *direct interventions* that exploit psychological weaknesses should also remain unregulated across the board.

More generally, showing that indirect interventions can produce equally normatively worrisome effects than direct ones is a strategy insufficient to validate the parity principle. The unfortunate existence of some influences does not justify retaining or putting into place further suspect influences. Consider this analogy. The inability to control all toxins in the air does not justify more emissions. Levy correctly writes that “people can manipulate others through [indirect interventions] just as effectively as they can using oxytocin” or other direct interventions (previous chapter). But what follows from this? I claim that

even though two interventions cause sufficiently similar effects, assessments may differ, frequently along the direct–indirect divide, because *other* normative considerations come into play—the many finer-grained considerations of the sort outlined in this chapter. This claim cannot be falsified by showing (dis)similarities of effects. Instead, it would have to be shown that other considerations are normatively *irrelevant*.

Taking Levy’s argument further, one may deny the significance of the direct–indirect distinction by holding that the by far strongest influences on everyone, inescapable and potent to shape deep levels of personalities, are upbringing, education, media, and culture. Indeed, cultural forces (and the industry creating them) shape our thoughts, desires, and even self-conceptions probably to a far greater degree than most direct interventions ever can. Nevertheless, although the right to mental self-determination is not blind to the social and cultural forces shaping everyone these forces largely escape meaningful legal regulation, at least as long as one seeks social participation. The right suggests scrutinizing particular cultural practices detrimental to mental self-determination (e.g., effects of photoshopped models on the self-image of adolescents). Stronger regulation seems warranted, and mental self-determination provides a strong argument to this end. However, the influence of such practices pales in comparison to the powers of broader social and cultural forces. The law is often impotent to capture, let alone regulate, dynamics and effects at these levels. But, nonetheless, given that everyone is exposed, even subjected to a range of such overwhelming influences—should therefore other (direct) interventions be permissible *as well*?

I don’t think so. I join Levy in lamenting social injustices such as lack of education and nutrition and their effects on mental development. I wish to add that the unequal distribution of means to change minds (e.g., the lacking access of millions of marginalized people to have their voices heard in the forums of political power) significantly contributes to this malaise (see Paulo & Bublitz, 2019). But I fail to see why a “concern for justice mandates ignoring the causal route whereby interventions work” (foregoing chapter). Society should not stop at banning direct interventions for wrongly assuming other forms of influence are ineffective, but begin by banning those that clearly undermine mental self-determination.

After all, as Levy seems to share the normative premise of mental self-determination, and as I agree with his descriptive claims, I can only account for our diverging conclusions by a final intuition: In a consequentialist framework, effects are all that matters *per definitionem* whereas rights-based frameworks usually insist on the importance of factors overriding consequences. It might not be a coincidence that a structurally similar contrast emerges here. Perhaps we are simply witnessing another variation of differences in more general positions of normative ethics.

Conclusion

In conclusion, if one seeks to secure an inner citadel from physiological or psychological influences, any of these proposed regulations are insufficient. But that aim is futile. Reasonable regulations for real minds start with accepting that the human mind is malleable and manipulable on diverse ways, many of which are hardly regulatable. Normatively, every person has a right against mind-interventions and, correlatively, a duty to respect mental self-determination of others. What this means more precisely is a complex question. Contemporarily, the ideal of mental self-determination is accorded an adequate role neither in legal or ethical theory, nor in practice. In its light, interventions into other minds that undermine or bypass control are particularly worrisome. Whether this is the case depends on the strength and mode of an intervention. Blanket prohibitions of all mind-altering stimuli, including sensory ones, are inconceivable in cooperative and communicative societies. For good reasons, particular forms of social interaction are considered desirable. This is reflected in specific rights to send mind-altering stimuli, primarily freedom of speech. Ultimately, we have to accept that social life necessarily involves influencing others and being influenced by others. Reasonable regulation has to pick out the most worrisome interventions. The distinction between physiological and psychological interventions is central and suggests a normative dualism: prima facie direct interventions are impermissible, whereas indirect interventions are permissible. Nonetheless, some forms of indirect interventions (e.g., akin to psychological trickery) may also be regarded as illicit. As I have hopefully urged enough in this chapter, finer distinctions with respect to particular contexts and means are necessary. The parity principle unfortunately gets in the way of seeing the necessity for systematically developing more nuanced distinctions.

References

- Andersson, M., Avaliani, N., Svensson, A., Wickham, J., Pinborg, L., Jespersen, B., . . . Kokaia, M. (2016). Optogenetic control of human neurons in organotypic brain cultures. *Scientific Reports*, 6, 24818.
- Baars, B. J. (1997). *In the theater of consciousness: The workspace of the mind*. New York, NY: Oxford University Press.
- Bargh, J. (2017). *Before you know it*. New York, NY: Touchstone.
- Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science*, 15(1), 1–4.
- Boire, R. G. (2001). On cognitive liberty. *The Journal of Cognitive Liberties*, 2(1), 7–22.
- Bublitz, C. (2014a). Cognitive liberty or the international human right to freedom of thought. In N. Levy & J. Clausen (Eds.), *Springer handbook of neuroethics* (pp. 1309–1333). Dordrecht, The Netherlands: Springer.
- Bublitz, C. (2014b). Freedom of thought in the age of neuroscience. *Archiv fuer Rechts- und Sozialphilosophie*, 100, 1–25.
- Bublitz, C. (2016). Drugs, enhancements & rights: Ten points for lawmakers to consider. In F. Jotterand & V. Dubljevic (Eds.), *Cognitive enhancement: Ethical and policy implications in international perspective* (pp. 309–328). New York: Oxford University Press.
- Bublitz, C. (2018). “The soul is the prison of the body”—Mandatory moral enhancement, punishment & rights against neuro-rehabilitation. In D. Birks & S. Thomas (Eds.), *Treatment for crime: Philosophical essays on neurointerventions in criminal justice* (pp. 289–320). Oxford, England: Oxford University Press.
- Bublitz, 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.
- Chalmers, D. (2002). Consciousness and its place in nature. In S. P. Stich & T. A. Warfield (Eds.), *The Blackwell guide to philosophy of mind* (pp. 102–142). Malden, MA: Blackwell.
- Chomsky, N. (1971). The case against BF Skinner. *The New York Review of Books*, 17(11), 18–24.
- Cialdini, R. (2007). *Influence: The psychology of persuasion*. New York, NY: HarperCollins.
- Cialdini, R. (2016). *Pre-suasion: A revolutionary way to influence and persuade*. London, England: Random House.
- Coons, C., & Weber, M. (2014). *Manipulation: Theory and practice*. Oxford, England: Oxford University Press.
- Crutchfield, P. (2016). The epistemology of moral bioenhancement. *Bioethics*, 30(6), 389–396.
- Dahaner, J. (2019). Why internal moral enhancement might be politically better than external moral enhancement. *Neuroethics*, 12(1), 39–54.
- Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., . . . Sheridan, J. F. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine*, 65(4), 564–570.
- Davis, N. J., van Koningsbruggen, M. G. (2013). “Non-invasive” brain stimulation is not non-invasive. *Frontiers in Systems Neuroscience*, 7, 76.
- Dijksterhuis, A., Aarts, H., & Smith, P. (2005). The power of the subliminal: On subliminal persuasion and other potential applications. In R. Hassin & J. Uleman (Eds.), *The new unconscious* (pp. 77–106). Oxford, England: Oxford University Press.
- Dougherty, D. D., Rezaei, A. R., Carpenter, L. L., Howland, R. H., Bhati, M. T., O’Reardon, J. P. & Cusin, C. (2015). A randomized sham-controlled trial of deep brain stimulation of the ventral capsule/ventral striatum for chronic treatment-resistant depression. *Biological Psychiatry*, 78(4), 240–248.
- Douglas, T. (2014). Criminal rehabilitation through medical intervention: Moral liability and the right to bodily integrity. *The Journal of Ethics*, 18, 101–122.
- Douglas, T. (2018). Neural and environmental modulation of motivation: What’s the moral difference? In D. Birks & T. Douglas (Eds.), *Treatment for crime: Philosophical essays on neurointerventions in criminal justice* (pp. 280–223). Oxford, England: Oxford University Press.

- Evans, J., & Frankish, K. (2009). *In two minds: Dual processes and beyond*. Oxford, England: Oxford University Press.
- Evans, J., & Stanovich, K. E. (2013). Dual process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, 8(3), 223–241.
- Focquaert, F., & Schermer, M. (2015). Moral enhancement: Do means matter morally? *Neuroethics*, 8(2), 139–151.
- Foucault, M. (1995). *Discipline and punish: The birth of the prison*. New York, NY: Second Vintage Books.
- Frankish, K. (2009). Systems and levels: Dual-system theories and the personal-subpersonal distinction. In J. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 89–107). Oxford, England: Oxford University Press.
- Genschow, O., Noll, T., Wänke, M., & Gersbach, R. (2015). Does Baker-Miller pink reduce aggression in prison detention cells? A critical empirical examination. *Psychology, Crime & Law*, 21(5), 482–489.
- Goode, E. (2016, July 29). Is hypnosis all in your head? Brain scans suggest otherwise. *The New York Times*. <http://www.nytimes.com/2016/07/30/science/hypnosis-brain-changes.html>
- Grassian, S., & Friedman, N. (1986). Effects of sensory deprivation in psychiatric seclusion and solitary confinement. *International Journal of Law and Psychiatry*, 8(1), 49–65.
- Greene, J. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, 124, 695–726.
- Hartogsohn, I. (2016). Set and setting, psychedelics and the placebo response: An extra-pharmacological perspective on psychopharmacology. *Journal of Psychopharmacology*, 30(12), 1259–1267.
- Hassin, R. (2005). Nonconscious control and implicit working memory. In R. Hassin & J. Uleman (Eds.), *The new unconscious* (pp. 196–223). Oxford, England: Oxford University Press.
- Jakubovski, E., Varigonda, A., Freemantle, N., Taylor, M., & Bloch, M. H. (2016). Systematic review and meta-analysis: Dose-response relationship of selective serotonin reuptake inhibitors in major depressive disorder. *American Journal of Psychiatry*, 173(2), 174–183.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Kahneman, D., & Shane, F. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York, NY: Cambridge University Press.
- Kandel, E. (2007). *In search of memory: The emergence of a new science of mind*. New York, NY: W. W. Norton.
- Kim, J. (2011). *Philosophy of mind*. Boulder, CO: Westview Press.
- Krohs, U., & Kroes, P. (2009). *Functions in biological and artificial worlds: Comparative philosophical perspectives*. Cambridge, MA: MIT Press.
- Leibniz, G. W. (2008). Letter to Friedrich Bierling. In C. I. Gerhardt (Ed.), *Die philosophischen Schriften von G. W. Leibniz* (Vol. 7). Hildesheim, Germany: Olms. (Original work published 1709)
- Lemma, A. (2016). *Introduction to the practice of psychoanalytic psychotherapy*. Malden, MA: Wiley.
- Levy, N. (2007). *Neuroethics: Challenges for the 21st century*. Cambridge, England: Cambridge University Press.
- Linden, D. (2006). How psychotherapy changes the brain—The contribution of functional neuroimaging. *Molecular Psychiatry*, 11, 528–538.
- Macpherson, F. (2011). Individuating the senses. In F. Macpherson, *The senses: Classical and contemporary philosophical perspectives* (pp. 3–45). New York, NY: Oxford University Press.
- Martin, E., Jeanmonod, D., Morel, A., Zadicario, E., & Werner, B. (2009). High-intensity focused ultrasound for noninvasive functional neurosurgery. *Annals of Neurology*, 66, 858–861. doi:10.1002/ana.21801
- Mayberg, H., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., . . . Kennedy, S. H. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, 45(5), 651–660.
- McLaughlin, P. (2000). *What functions explain: Functional explanation and self-reproducing systems*. Cambridge, England: Cambridge University Press.
- Merkel, R. (2007). Treatment—Prevention—Enhancement: Normative foundations and limits. In R. Merkel, G. Boer, J. Fegert, & T. Galert (Eds.), *Intervening in the brain: Changing psyche and society* (pp. 289–382). Heidelberg, Germany: Springer.
- Merriam-Webster. (n.d.). Medical definition of *invasive*. Retrieved from <https://www.merriam-webster.com/dictionary/invasive>
- Metzinger, T. (2010). *The ego tunnel: The science of the mind and the myth of the self*. New York, NY: Basic Books.
- Metzinger, T. (2013). The myth of cognitive agency: Subpersonal thinking as a cyclically recurring loss of mental autonomy. *Frontiers Psychology*, 4, 931.
- Morishita, T., Fayad, S., Higuchi, M., Nestor, K. A., & Foote, K. D. (2014). Deep brain stimulation for treatment-resistant depression: Systematic review of clinical outcomes. *Neurotherapeutics*, 11(3), 475–484. doi:10.1007/s13311-014-0282-1
- Nussbaum, M. C. (1995). Objectification. *Philosophy & Public Affairs*, 24(4), 249–291.
- Paulo, N., & Bublitz, C. (2019). Pow(d)er to the people? Voter manipulation, legitimacy, and the relevance of moral psychology for democratic theory. *Neuroethics*, 12, 55–71.
- Prinz, J. (2016). Unconscious perception. In M. Matthen (Ed.), *The Oxford handbook of philosophy of perception* (pp. 371–389). Oxford, England: Oxford University Press.
- Quiroga, R. Q. (2012). Concept cells: The building blocks of declarative memory functions. *Nature Reviews Neuroscience*, 13(8), 587–597.
- Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045), 1102–1107.
- Reti, I. (Ed.). (2015). *Brain stimulation: Methodologies and interventions*. Hoboken, NJ: Wiley.
- Schlaepfer, T., & Lieb, K. (2005). Deep brain stimulation for treatment of refractory depression. *The Lancet*, 366(9495), 1420–1422.
- Scholey, A., & Owen, L. (2013). Effects of chocolate on cognitive function and mood: A systematic review. *Nutrition Reviews*, 71(10), 665–681. doi:10.1111/nure.12065
- Seung, S. (2012). *Connectome: How the brain's wiring makes us who we are*. Boston, MA: Houghton Mifflin Harcourt.
- Skinner, B. F. (2002). *Beyond freedom and dignity*. London: Hackett.
- Specker, J., Schermer, M., & Reiner, P. (2017). Public attitudes toward moral enhancement: Evidence that means matter morally. *Neuroethics*, 10, 405–417.
- Stokes, D. (2012). Perceiving and desiring: A new look at the cognitive penetrability of experience. *Philosophical Studies*, 158(3), 477–492.
- Sunstein, C. (2016). Do people like nudges? *Administrative Law Review*, 68(2), 177–232.

- Sunstein, C., & Thaler, R. (2003). Libertarian paternalism is not an oxymoron. *University of Chicago Law Review*, 70(4), 1159–1202.
- Taylor, K. (2006). *Brainwashing: The science of thought control*. Oxford, England: Oxford University Press.
- Thaler, R., & Sunstein, C. (2008). *Nudge: Improving decisions about health, wealth and happiness*. New Haven, CT: Yale University Press.
- Tomasello, M. (2016). *A natural history of human morality*. Cambridge, MA: Harvard University Press.
- Valenstein, E. (1973). *Brain control: A critical examination of brain stimulation and surgery*. New York, NY: Wiley.
- Vina, J., Sanchis-Gomar, F., Martinez-Bello, V., & Gomez-Cabrera, M. C. (2012). Exercise acts as a drug: The pharmacological benefits of exercise. *British Journal of Pharmacology*, 167, 1–12.
- Wegner, D. (2005). Who is the controller of controlled processes? In R. Hassin & J. Uleman (Eds.), *The new unconscious* (pp. 19–36). Oxford, England: Oxford University Press.
- Wenk, G. (2015). *Your brain on food: How chemicals control your thoughts and feelings*. Oxford, England: Oxford University Press.
- Wheeler, H. (Ed.). (1973). *Beyond the punitive society*. San Francisco, CA: W. H. Freeman.
- Zawidzki, T. (2013). *Mindhaping: A new framework for understanding social cognition*. Cambridge, MA: MIT Press.
- Zinberg, N. (1984). *Drug, set, and setting: The basis for controlled intoxicant use*. New Haven, CT: Yale University Press.

4

Neuroprosthetics, Behavior Control, and Criminal Responsibility

Walter Glannon

Introduction

Philosophers and legal theorists argue that moral and criminal responsibility for actions require autonomous agency. The bodily movements with which the actions are identified must be voluntary and must issue from mental states the person endorses following critical reflection. In addition, the action must not involve any physical or psychological coercion, compulsion, or constraint (Dworkin, 1988; Frankfurt, 1988a, 1988b, 1988c; Mele, 1995). Many also argue that moral and legal responsibility for actions presuppose that they do not result from causal routes that bypass the agent's mental states as the direct causes of her actions (Davidson, 2001a, 2001b; Mele, 1995, 2009). Agents have to act from their "own mechanisms, which cannot be formed by pills, *electronic stimulation* of the brain or brainwashing" (Fischer & Ravizza, 1998, p. 236, emphasis added; also see Bublit & Merkel, 2013). Being morally responsible for one's actions excludes "severe manipulation of the brain, hypnosis and the like" (Fischer, 2006, p. 53). These conditions would seem to undermine responsiveness to reasons as a necessary condition for responsibility.

If the previous discussion is right, however, then it may seem that in bypassing, replacing, or modulating damaged neural circuits mediating motor and mental functions, neuroprosthetics undermine the conscious control necessary for autonomous and responsible agency. Mechanisms other than the person's own normally functioning brain-mind seem to undermine the control necessary for responsibility because they seem to cause nonvoluntary or involuntary actions. Yet when they operate effectively, neuroprosthetics surely do not undermine but restore control of motor and mental functions when they have been lost through brain injury or impaired by neurodevelopmental or neurodegenerative disorders (Glannon, 2015; Schermer, 2015; Vincent, 2015). They enable varying degrees of voluntary agency and responsibility by restoring varying degrees of the requisite motor and mental capacities. For this reason, whether or to what extent a person with a device implanted in his brain can be criminally responsible for an action,