New Waves in Philosophy
Series Editors: Vincent F. Hendricks and Duncan Pritchard

Titles include:
Jesús H. Aguilar, Andrei A. Buckareff and Keith Frankish (*editors*)
NEW WAVES IN PHILOSOPHY OF ACTION

Michael Brady
NEW WAVES IN METAETHICS

Otavio Bueno and Oystein Linnebo (*editors*)
NEW WAVES IN PHILOSOPHY OF MATHEMATICS

Boudewijn DeBruin and Christopher F. Zurn (*editors*)
NEW WAVES IN POLITICAL PHILOSOPHY

Allan Hazlett (*editor*)
NEW WAVES IN METAPHYSICS

Vincent F. Hendricks and Duncan Pritchard (*editors*)
NEW WAVES IN EPISTEMOLOGY

P.D. Magnus and Jacob Busch (*editors*)
NEW WAVES IN PHILOSOPHY OF SCIENCE

Yujin Nagasawa and Erik J. Wielenberg (*editors*)
NEW WAVES IN PHILOSOPHY OF RELIGION

Jan Kyrre Berg Olsen, Evan Selinger and Søren Riis (*editors*)
NEW WAVES IN PHILOSOPHY OF TECHNOLOGY

Thomas S. Petersen, Jesper Ryberg and Clark Wolf (*editors*)
NEW WAVES IN APPLIED ETHICS

Sarah Sawyer (*editor*)
NEW WAVES IN PHILOSOPHY OF LANGUAGE

Kathleen Stock and Katherine Thomson-Jones (*editors*)
NEW WAVES IN AESTHETICS

Cory D. Wright and Nikolaj J. L. L. Pedersen (*editors*)
NEW WAVES IN TRUTH

Forthcoming:
Thom Brooks (*editor*)
NEW WAVES IN ETHICS

Maksymilian Del Mar
NEW WAVES IN PHILOSOPHY OF LAW
New Waves in Philosophy
(outside North America only)

You can receive future titles in this series as they are published by placing a standing order. Please contact your bookseller or, in case of difficulty, write to us at the address below with your name and address, the title of the series and the ISBN quoted above.

Customer Services Department, Macmillan Distribution Ltd, Houndmills, Basingstoke, Hampshire RG21 6XS, England
New Waves in Philosophy of Action

Edited by

Jesús H. Aguilar
*Rochester Institute of Technology, USA*

Andrei A. Buckareff
*Marist College, USA*

Keith Frankish
*The Open University, UK*
## Contents

**List of Figures** vii

**Series Editors’ Preface** viii

**Notes on Contributors** ix

**Acknowledgments** xi

**Introduction** 1

*Jesús Aguilar, Andrei Buckareff, and Keith Frankish*

### Part I Foundations of Action and Agency 11

1 **Agency, Ownership, and the Standard Theory** 13
   *Markus E. Schlosser*

2 **Failing to Do the Impossible** 32
   *Carolina Sartorio*

3 **Experimental Philosophy of Action** 50
   *Thomas Nadelhoffer*

### Part II Autonomy 79

4 **Identification, Psychology, and Habits** 81
   *Bill Pollard*

5 **Mass Perverse Identification: An Argument for a Modest Concept of Autonomy** 98
   *Yonatan Shemmer*

6 **Cartesian Reflections on the Autonomy of the Mental** 122
   *Matthew Soteriou*

### Part III Free Agency 141

7 **The Revisionist Turn: A Brief History of Recent Work on Free Will** 143
   *Manuel Vargas*

8 **Luck and Free Will** 173
   *Neil Levy*

9 **Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions** 189
   *Eddy Nahmias and Dylan Murray*
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>Agentive Experiences as Pushmi-Pullyu Representations</td>
<td>219</td>
</tr>
<tr>
<td></td>
<td>Tim Bayne</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Double Bookkeeping in Delusions: Explaining the Gap</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>between Saying and Doing</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lisa Bortolotti</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>The Limits of Rationality in Collective Action Explanations</td>
<td>257</td>
</tr>
<tr>
<td></td>
<td>Sara Rachel Chant</td>
<td></td>
</tr>
</tbody>
</table>

*Index* 274
List of Figures

9.1 *Genuine* incompatibilist intuitions 193
9.2 *Apparent* incompatibilist intuitions 193
9.3 Judgments about MR, FW, and bypassing 203
9.4 Apparent incompatibilists who are bypassers and prima facie compatibilists who are not 204
9.5 MR/FW and bypassing descriptive statistics 205
9.6 Mediation analysis 206
Series Editors’ Preface

New Waves in Philosophy Series

The aim of this series is to gather the young and up-and-coming scholars in philosophy to give their view of the subject now and in the years to come, and to serve a documentary purpose, that is, “this is what they said then, and this is what happened.” It will also provide a snapshot of cutting-edge research that will be of vital interested to researchers and students working in all subject areas of philosophy.

The goal of the series is to have a New Waves volume in every one of the main areas of philosophy. We would like to thank Palgrave Macmillan for taking on this project in particular, and the entire New Waves in Philosophy series in general.

Vincent F. Hendricks and Duncan Pritchard
Notes on Contributors

Tim Bayne is University Lecturer in the Philosophy of Mind at the University of Oxford and a Fellow of St. Catherine's College. He is the author of a number of articles on consciousness, and is an editor of The Oxford Companion to Consciousness. His book The Unity of Consciousness is shortly to be published by Oxford University Press.

Lisa Bortolotti is Senior Lecturer in Philosophy at the University of Birmingham (UK). Her main research interests are in the philosophy of the cognitive sciences, with a particular focus on irrationality, self-knowledge, and agency. She is the author of Delusions and Other Irrational Beliefs (Oxford University Press, 2009) and the editor of Philosophy and Happiness (Palgrave Macmillan, 2009). She also writes about biomedical ethics, especially in the context of scientific research and reproduction.

Sara Rachel Chant is Assistant Professor of Philosophy at the University of Missouri-Columbia. Her main areas of research include metaphysics and philosophy of action, with special concern paid to the methodological approach taken towards classical problems in these areas. She has published her work in journals such as Pacific Philosophical Quarterly, Philosophical Explorations, Studia Logica, and Philosophical Studies. Chant is currently completing a book with Zachary Ernst on the epistemology of collectivity.

Neil Levy is Director of Research at the Oxford Centre for Neuroethics and Head of Neuroethics at the Florey Neuroscience Institutes (Melbourne, Australia). He is the author of five books, including Neuroethics (Cambridge University Press, 2007) and around 100 articles in applied ethics, philosophical psychology, and free will and moral responsibility. He is currently writing a book on luck and free will.

Dylan Murray is a PhD student in the Department of Philosophy at the University of California, Berkeley. He was previously a Brains & Behavior Fellow in the Philosophy Department and Neuroscience Institute at Georgia State University. His research focuses on action theory, moral psychology, and philosophy of mind.

Thomas Nadelhoffer is Assistant Professor of Philosophy at Dickinson College. His research focuses on action theory, moral psychology, and the philosophy of law. He has published his work in journals such as Analysis, Midwest Studies in Philosophy, Mind & Language, and Philosophy and Phenomenological Research.
Eddy Nahmias is Associate Professor in the Philosophy Department and the Neuroscience Institute at Georgia State University. He specializes in philosophy of mind/cognitive science, free will, moral psychology, and experimental philosophy. He has published numerous chapters and articles in these areas, co-editor of *Moral Psychology: Classic and Contemporary Readings* (Wiley-Blackwell), and is writing *Rediscovering Free Will* (Oxford University Press), which examines potential threats to free will from research in the mind sciences.

Bill Pollard is an Honorary Research Fellow at the University of Edinburgh. He has taught at the universities of Edinburgh, Warwick and York in the UK, and Dartmouth College in the USA. His main research interests are in the philosophy of action, with particular focus on the habits, their rationality and explanatory role. He is the author of *Habits in Action* (VDM Verlag Dr. Mueller e.K., 2008) and a number of articles on habits.

Carolina Sartorio is Associate Professor of Philosophy at the University of Arizona. Her main research interests are in metaphysics, ethics, and their intersection. She has published articles on the topics of causation, moral responsibility, agency, and free will.

Markus Schlosser is a Research Fellow at the University of Leiden. His main research area is the philosophy of action, with a focus on metaphysical issues. He has published his work in journals such as *Analysis*, and *Philosophical Explorations*. Currently, he is working on a philosophical evaluation of empirical claims concerning mental causation and free will.

Yonatan Shemmer is a Lecturer of Philosophy at The University of Sheffield, UK. His main interests are in practical reason and normative psychology. He has recently spent time thinking about constructivist views of practical reason and is co-editing a volume on Constructivism in practical philosophy. He is the author of “Desires as Reasons” (2007) as well as some other articles on Humeanism in practical reason.

Matthew Soteriou is Reader in Philosophy at the University of Warwick. His main research interests are in the philosophy of mind and action, and epistemology, with a particular focus on perception, mental action and the ontology of mind.

Manuel Vargas is Associate Professor of Philosophy at the University of San Francisco. His main areas of interest include the philosophy of agency, moral psychology, and Latin American philosophy. He is one of the authors (along with John Martin Fischer, Robert Kane, and Derk Pereboom) of *Four Views on Free Will* (Blackwell, 2007), and has written articles on such topics as free will, practical reason, evil, Eurocentrism, and the undead. He is currently working on a book on moral responsibility.
Acknowledgments

A volume such as this is the result of the hard work of a number of people. The editors would first like to thank the authors of the essays for their exceptional contributions to this volume. A debt of gratitude is also due to the editors of the *New Waves in Philosophy* series, Vincent F. Hendricks and Duncan Pritchard, for their support of this project. We are also grateful to Evan Selinger for putting us into contact with Vincent and encouraging us to contribute a volume to the series. Our contacts at Palgrave Macmillan, Priyanka Gibbons and Melanie Blair, deserve special thanks for their patience with us and willingness to be accommodating when we requested extra time to work on the manuscript. From Macmillan Publishing Solution Limited, we would like to thank Imran Shahnawaz and his production team for their editorial work.

The task of putting this volume together was greatly facilitated by the work of the many referees who were willing to review the articles for us. These included Kent Bach, John Bishop, Ish Haji, Bennett Helm, Terry Horgan, Ron Mallon, Michael McKenna, Alfred R. Mele, Shaun Nichols, Lucy O’Brien, Joëlle Proust, and Christine Tappolet. We are grateful to them all for their hard work.

We also wish to thank our home institutions for support. In particular, Andrei is grateful to Marist College for giving him a course release that allowed him extra time to work on this and other projects. Jesús is very thankful for the support received from the Rochester Institute of Technology in the form of course releases as the recipient of the Paul A. and Francena L. Miller Faculty Fellowship. Keith is grateful to The Open University for support during some difficult personal times, and to the Department of Philosophy and Social Studies at the University of Crete for welcoming him as a Visiting Researcher. In particular, he would like to thank Maria Venieri for helping him feel part of the philosophy community in Crete.

Finally, all three editors are especially grateful to their loved ones. Jesús wishes to thank Amy Wolf for her unfaltering support and loving understanding during his work on this volume. Andrei is grateful to his spouse, Lara E. Kasper-Buckareff for her encouragement, loving support, and legal advice. Keith wishes to thank his wife, Maria Kasmirli, for her patience, support, and excellent advice on philosophical matters.
Introduction

Jesús Aguilar, Andrei Buckareff, and Keith Frankish

1 Philosophy of action

This volume contains a set of cutting-edge essays by younger philosophers on various topics in what can broadly be characterized as the philosophy of action. Some of the essays are about the metaphysics of action and agency; some consider the nature of autonomy and free agency; some explore conceptual and normative issues; some draw on data from psychology and psychopathology. But what all of them have in common is that they address some problem related to our existence as human agents.

The range of topics covered in this collection is broad. This is intentional. Rather than focus on one narrow topic, we have chosen to collect papers that, taken together, introduce readers to some key debates in contemporary philosophy of action. Of course, exactly how broad a range of issues one will expect to find in such a volume will depend on how one characterizes the field of philosophy of action, and we shall begin by saying something about how we characterize it.

So what is the philosophy of action? One view treats it as either a narrow sub-discipline of philosophy or as a proper part of another sub-discipline or set of sub-disciplines, primarily metaphysics and the philosophy of mind. On this view, philosophy of action is focused on the conceptual, epistemological, and metaphysical issues surrounding the nature of action and its explanation (action-individuation, the nature of intentions and their role in intentional action, the nature of the will, reasons and their role in explaining action, and so on).

A problem with this characterization is that it winds up separating a host of closely related issues, excluding certain problems (for example, akrasia and enkrasia, free will, motivation) from the field simply because they do not concern foundational issues about the nature and explanation of human action. To illustrate what is wrong with this approach, we may make a comparison with epistemology. Much of the work on topics such as epistemic akrasia, epistemic virtue, and reasonable disagreements does
not bear directly on questions about the nature of knowledge itself. (For example, one can theorize about the epistemic virtues but endorse a theory of knowledge that dispenses with any role whatsoever for epistemic virtues.) Do these topics belong outside epistemology? Since they all bear on issues related to the broader interests of epistemologists, including the nature of epistemic evaluations, they belong in epistemology. Something similar is the case with respect to the philosophy of action, where it is hard to make a clean separation between foundational issues about intentional action and wider issues about, for example, the nature of free agency. Indeed, one’s view about a foundational issue may directly affect how one thinks about a non-foundational issue, and vice versa.

Another way of characterizing the philosophy of action is as a broad sub-discipline of philosophy. On this view, which is the one taken in this volume, the philosophy of action is a wide field that overlaps with other areas in important ways, but has its own distinctive set of questions and problems, all of which are broadly related to questions about the nature, explanation, and scope of our agency and action. So conceived, philosophy of action embraces not only traditional concerns about the nature and explanation of intentional action, but also a host of other issues, including, among others, the problems of akrasia and enkrasia, free will and moral responsibility, mental agency, motivation, omissions, practical reason, and social action. Of course, philosophy of action is closely related to other philosophical sub-disciplines. For instance, philosophy of mind and philosophy of action both address questions about mental action, mental causation, and reasons-explanations. This should not be surprising, however. As in other areas of philosophy, the boundaries between the various sub-disciplines are vague, relating to each other like the intersecting circles of a Venn diagram.

One objection to this view of the field is that the philosophy of action lies at the intersection of so many disparate sub-disciplines that there is very little glue to hold it together, so to speak. After all, in the course of their career, a philosopher may become an expert on the ontological issues in the philosophy of action without developing the same sophisticated understanding of the normative debates (which require a level of expertise in ethics that an ontologist may lack). Indeed, very few philosophers (Alfred Mele being a rare example) have proved able to move with ease among the various topics gathered under the banner of philosophy of action. So perhaps we should reject the idea that there is a unified sub-discipline identifiable as the philosophy of action. Rather, there are just problems about action and agency in the myriad sub-disciplines of philosophy that bear a family resemblance to each other. And—the objection goes—such a family resemblance is not enough to constitute the philosophy of action as a distinct sub-discipline in its own right.

However, this objection ignores the fact that, once again, philosophy of action is not unique in this respect. Compare the philosophy of mind.
Here, too, a philosopher may spend most of their career focusing on a narrow set of issues that do not bear directly on debates elsewhere in the field. One researcher might work on the problem of mental causation while another focuses almost entirely on the nature of phenomenal consciousness. Although there is some overlap between these two areas, the two researchers would need to undertake substantial retraining and scholarship if they suddenly decided to swap topics. The point is that philosophy of mind is not defined by a core body of knowledge and expertise shared by all practitioners. Rather, it is constituted by a set of linked problems, to which different bodies of knowledge and expertise need to be applied. In short, philosophy of mind, too, is unified only by a family resemblance among the various problems it encompasses—and is none the worse for that.

An advantage of characterizing the philosophy of action as a broad sub-discipline is that it puts pressure on philosophers of action to pay attention to work done in other areas of the field, which, they may find, bears directly on their own. To give just one example, if one concludes that there is nothing more to an agent’s exercise of executive control than that the right type of mental states non-deviantly cause their behavior, then one may find that certain theories of autonomous control and free agency are no longer viable options. Of course, philosophers of action also need to pay attention to work in other philosophical sub-disciplines and in the relevant sciences. Indeed, recent years have seen increasing collaboration between philosophers, psychologists, and neuroscientists with respect to the study of agency and action, and many philosophers have made it a point to do work that is empirically informed (Myles Brand, Alfred Mele, Fred Adams, and Timothy O’Connor all come to mind). But the importance of such interdisciplinary connections does not weaken the case for treating philosophy of action as a broad sub-discipline within philosophy—the opposite, if anything.

Of course, in this short introduction we can only sketch the case for our view of the philosophy of action, and some readers may be unconvinced. Still, even if the philosophy of action is not an independent sub-discipline, there is no doubt that there are problems about human action and agency to be found in all of the major sub-disciplines of philosophy. The essays in this volume address many of these problems. And, though the coverage is not exhaustive, readers new to the field should come away from the volume with a good sense of current thinking about human action and agency. For their part, established researchers in the field will, we believe, find the essays to be original contributions that substantially advance many debates about action and agency.

2 The essays

The essays in this volume are divided between four parts. Those in the first part address issues at the foundation of theorizing about action and agency.
Those in the second and third parts relate to autonomy and free agency respectively, and those in the final part address issues which in various ways extend the boundaries of traditional philosophy of action.

2.1 Foundations of action and agency

The essays in Part I look at foundational issues in the philosophy of action, using both traditional methods and new approaches.

In “Agency, Ownership, and the Standard Theory,” Markus Schlosser defends the Causal Theory of Action (CTA), according to which an agent-involving event counts as an action if it is caused in the right way by appropriate agent-involving mental states. The CTA is considered the standard theory of action, and Schlosser shows why it has retained this status, presenting its main features and arguing that it offers the only viable account of the metaphysics of agency. A major challenge to the CTA is that it is unable to provide an account of the role of the agent. On one version of the objection, the CTA reduces activity to mere happenings, and thus fails to capture the phenomenon of agency. On another, weaker, version of the objection, the CTA fails to capture the role played by the human agent in certain higher forms of action. In response, Schlosser argues that the CTA does support a robust notion of agency, especially when fleshed out with a feedback-comparator model of motor control, and that higher kinds of agency can be accounted for by supplementing the CTA in various ways.

In her contribution, “Failing to do the Impossible,” Carolina Sartorio considers the relationship between intentional omissions and alternative possibilities. Specifically, she considers whether an agent’s intentionally omitting to do something requires that the agent could have intentionally acted otherwise. Engaging with related work on moral responsibility and on omissions (including Harry Frankfurt’s work on both), she provides an account of intentional omissions that can handle challenging cases that evoke competing intuitions. She contends that an agent can intentionally omit to act even if it is impossible for them to have acted otherwise. While this result may be welcome to those who contend that morally responsible agency does not require the ability to act contrary to how one actually acted, the same people may not welcome Sartorio’s conclusion that there is an important asymmetry between the conditions for intentionally acting and for intentionally omitting to act.

In the final chapter in Part I, “Experimental Philosophy of Action,” Thomas Nadelhoffer surveys work in the experimental philosophy of action. Many philosophers make claims about our everyday “folk” intuitions and judgments, and seek to develop theories that are broadly compatible with them. However, they have traditionally made these claims from the comfort of their armchairs, assuming that their own intuitions coincide with those of folk in general. Philosophers working in the new and rapidly developing field of experimental philosophy aim to correct this by
conducting systematic surveys of people’s intuitions on matters relevant to philosophical problems. This experimental approach has been particularly influential within philosophy of action, and Nadelhoffer argues for its importance in exploring the nature of intentional action. He focuses on Joshua Knobe’s “side effect” findings, which indicate that our judgments about whether or not an action is intentional are influenced by our moral evaluation of the action’s outcome—a view that runs counter to most traditional philosophical accounts of intentional action. Nadelhoffer reviews the key models that have been proposed to explain Knobe’s data and introduces some recent work that may shape the debate in the future. His overall aim is to provide the reader with a framework for understanding the field of experimental philosophy of action and encouragement to contribute to it.

2.2 Autonomy

In the recent literature on the philosophy of action, a distinction has arisen between two different ways in which agents exercise control over their actions, which François Schroeter has christened basic executive control and autonomous control respectively (Schroeter, 2004). Basic executive control is manifested whenever an agent exercises whatever sort of guidance is necessary to ensure the successful execution of their actions, and it is arguable that when we theorize about run-of-the-mill intentional action, we are concerned with basic executive control. Autonomous control, on the other hand, is a species of control that manifests our capacity for self-governance (“autonomous” literally means self-governed), and many writers claim that basic executive control, while necessary for autonomous control, is not sufficient for it. If that is right, then we shall need to supplement our theories of intentional action in order to get an account of autonomous control and, hence, of what J. David Velleman calls “full-blooded-action” (Velleman, 1992).

One of the most popular strategies here is to appeal to “endorsement” or “identification” as the necessary addition to executive control that will deliver autonomous control. According to this view, variants of which have been offered by, among others, Michael Bratman, Harry Frankfurt, and J. David Velleman, an agent exercises autonomous control over an action if the action issues from motivational states with which they identify or which they have endorsed. Although it is not the only strategy for accounting for autonomy, the identification strategy is one of the most popular. Still, there are problems with it, and other accounts of autonomy have been proposed, including ones that take basic executive control to be sufficient for autonomous control. The essays in Part II of the volume adopt varying approaches to autonomy while also exploring other issues.

The first essay, Bill Pollard’s “Identification, Psychology and Habits,” takes up the search for a theory of identification, which (following Frankfurt) he understands as an account of the difference between motives that are
external to the agent, such as drug cravings or sudden emotional impulses, and internal motives, with which the agent is identified. Pollard notes that existing accounts assume that for a motive to be internal it is necessary that it be suitably related to aspects of the agent’s psychology—whether psychological states, such as higher-order desires, or psychological actions, such as decisions or judgments. Pollard rejects these accounts, arguing that they are subject to a regress, and proposes that for a motive to be internal it is sufficient that it expresses one of the agent’s habits. If this is correct, then a theory of identification need not make reference to the agent’s psychological states at all. Pollard also offers a definition of “habit” and defends a conception of agency according to which an agent is partly constituted by their habits.

Yonatan Shemmer’s chapter “Mass Perverse Identification: An Argument for a Modest Concept of Autonomy” continues the themes of identification and autonomy. Shemmer notes that many philosophers subscribe to the so-called dependence thesis, according to which an agent performs an action autonomously only if they identify with the desire that motivates it. However, as Shemmer stresses, we very often act in ways that are not in accord with our value judgments—failing to donate to charity, shunning voluntary work, and so on. In such cases, he argues, we do not identify with the desires that motivate our actions, but the actions are autonomous all the same. Shemmer concludes that we must challenge the dependence thesis and accept that an action may be autonomous even if the agent does not identify with it. An action is autonomous, Shemmer proposes, provided it is part of a life that is composed of actions most of which the agent identifies with. Thus, ascriptions of autonomy, whether to agents or their actions, are always sensitive to global considerations. Shemmer defends this proposal from objections and argues that it comports with our core intuitions.

In his chapter “Cartesian Reflections on the Autonomy of the Mental,” Matthew Soteriou turns to the role of mental agency in autonomy. Taking his lead from remarks in Descartes’s Meditations, Soteriou focuses on two central but relatively neglected aspects of conscious thinking: supposition, in which a proposition that is not believed is temporarily adopted as a premise for the sake of argument, and bracketing, in which a proposition that is believed is temporarily excluded from use as a premise for the purpose of its epistemic evaluation. These acts, Soteriou argues, are not one-off events, but function as self-imposed constraints on our future reasoning, and in acknowledging and respecting them we manifest a distinctively human form of agency and autonomy.

2.3 Free agency

Part III of the volume is devoted to one of most prominent debates in the philosophy of action, and indeed in philosophy generally: the debate about free agency. There has been a resurgence of interest in this topic in recent years,
which has contributed to a growth of interest in a range of related issues in philosophy of action. In many respects, debates over free agency have proved more tractable than was previously expected, with concessions being made by various parties. What has happened, especially since the 1980s, is that there has been a shift in emphasis from the debate over the question of the compatibility of free will with determinism (although this is still prominent) to debates over competing models of free agency. At the same time, a stronger sense is emerging of how issues about the explanation of action and questions about the nature of free agency hang together, even if not always explicitly acknowledged in the literature.

In the opening essay of Part III, “The Revisionist Turn: A Brief History of Recent Work on Free Will,” Manuel Vargas focuses on the recent history of the free will debate, locating it within the context of metaphilosophical debates among analytic philosophers regarding the methods and goals of philosophical inquiry. In particular, Vargas examines the differences between those who see the free will problem as primarily a metaphysical issue and those whose work on the problem is motivated primarily by ethical concerns. The metaphysicians’ approach has elevated the role of intuitions in theorizing about free agency, whereas the ethicists have been more willing to adopt theories that are counter-intuitive but preferable on other grounds. Vargas sides with the ethicists, arguing that intuitions are a source of data when theorizing about free agency, but are not sacrosanct. Finally, as a way of highlighting some of the methodological challenges and options in theorizing about free agency, Vargas presents his own revisionist, responsibility-centric theory of free agency. He urges greater attention to the methodological matters he takes up in his essay in future theorizing about free will and moral responsibility.

In his essay, “Luck and Free Will,” Neil Levy focuses on the role of the notion of luck in the free will debate. Traditionally, the notion has been invoked in an argument against the libertarian view that free actions are undetermined. It is argued that an undetermined event is a matter of luck, and that actions that are a matter of luck do not manifest control or, consequently, freedom, on the part of the agent. Unlike many working in the field, who treat luck as a primitive notion or provide anemic accounts of it, Levy develops a robust analysis of luck, and he goes on to argue that it poses a challenge for all theories of free agency, be they libertarian or compatibilist.

Another recent development in the free will debate is the application of the methods of experimental philosophy, discussed previously in Thomas Nadelhoffer’s chapter. In their essay for this volume, “Experimental Philosophy on Free Will: An Error Theory for Incompatibilist Intuitions,” Eddy Nahmias and Dylan Murray survey recent work in this area and present a new study of their own. Their results point to an error theory for incompatibilist intuitions among ordinary folks. Laypeople, they argue,
confuse causal determinism with fatalism. In particular, most people who think causal determinism is incompatible with free will do so because they think it involves bypassing the means whereby agents exercise control. People who correctly understand causal determinism tend to regard it as compatible with free will and moral responsibility. This obviously poses a challenge for incompatibilists who hold that the folk are natural incompatibilists.

2.4 Action and agency in context

This final part includes three essays which in different ways push the boundaries of traditional philosophy of action and illustrate how issues of action and agency link up with wider concerns in philosophy of mind and cognitive science.

Tim Bayne’s paper “Agentive Experiences as Pushmi-pullyu Representations,” looks at the experience of agency—a topic that was, until recently, relatively neglected by both philosophers of action and philosophers of mind. In particular, Bayne asks how we should characterize the intentional structure and function of agentive experience. There are two established analyses here, thetic and telic. On the thetic analysis agentive experiences are descriptive, like beliefs; they represent how things are, and serve to track states of the world. On the telic analysis they are directive, like desires and intentions; they represent how we want things to be, and serve to direct action. Bayne examines both analyses before going on to propose a third, hybrid account, according to which agentive experiences are “pushmi-pullyu” representations, which have both thetic and telic structure. Bayne defends this view on empirical grounds, arguing that it best captures our current understanding of the cognitive role of agentive experiences.

In her paper, “Double Bookkeeping in Delusions: Explaining the Gap Between Saying and Doing,” Lisa Bortolotti considers the role of psychotic delusions in the guidance of action—an issue that lies at the intersection of philosophy of action and the rapidly developing field of philosophy of psychopathology. The standard view of delusions is that they are instances of belief—a view known as the doxastic account. However, delusions sometimes fail to guide action in the way that beliefs do (a phenomenon referred to as “double bookkeeping”), and this suggests that their contents are not objects of genuine belief. Bortolotti defends the doxastic account, arguing, first, that many delusions do guide action, at least some of the time; second, that ordinary beliefs, too, sometimes fail to guide action; and, third, that failures to act upon delusions may be due to lack of motivation rather than to lack of commitment to their truth. Thus Bortolotti emphasizes the continuities between normal and abnormal cognition, shedding light on the roles of both belief and delusion.

The final paper in the volume, Sara Rachel Chant’s “The Limits of Rationality in Collective Action Explanations,” explores the nature of actions
performed by more than one agent. Analyses of such actions traditionally proceed by identifying a set of reasons possessed by each of the participating individuals which justify their taking part in the collective action. Chant agrees that many cases fit this model. However, she also identifies a set of “difficult cases” where, for structural reasons, no such rational choice explanation is available. She explores these cases in detail and suggests that, in order to explain them, we may have to accept that people have an innate propensity to cooperate in certain circumstances. In studying some kinds of collective action, Chant concludes, we may need to adopt a similar approach to that taken in research on anomalous cooperative and altruistic behavior, where rational choice explanations are likewise unavailable.

3 Conclusion

As noted earlier, we do not pretend to have collected a group of essays that addresses all the issues in the philosophy of action. In fact, there are prominent perennial debates that are not taken up in any of the essays collected here. To attempt to cover every important debate would require a much larger volume, and one that would not obviously do what this collection is meant to do—namely expose readers to the work of some younger, up-and-coming philosophers of action on both ancient debates and new ones. The contributors to this volume will be leaders in the field for years to come. And this volume will have served its purpose if it allows both those new to the philosophy of action and seasoned veterans in the field to become acquainted with the work of these exceptional philosophers.

References

Part I
Foundations of Action and Agency
Agency, Ownership, and the Standard Theory

Markus E. Schlosser

The causal theory of action has been the standard view in the philosophy of action and mind. In the philosophy of mind, it is a piece of orthodoxy that is widely taken for granted and hardly ever questioned. In the philosophy of action, it has always had its critics. In this chapter, I will present responses to two challenges to the theory. The first says, basically, that there is no positive argument in favour of the causal theory, as the only reason that supports it consists in the apparent lack of tenable alternatives. The second challenge says that the theory fails to capture the phenomenon of agency, as it reduces activity to mere happenings (events and event-causal processes). This is often referred to as the problem of disappearing agency. A full defence of the causal theory should address both challenges. In the first part of this chapter, I will present what I take to be the core of the causal theory. In the second and the third part, I will then offer my responses to the two challenges. I will present a positive argument for the causal theory on the basis of considerations concerning the metaphysics of agency, and I will suggest that we own the agency that springs from our mental states and events by default.

My main aim is to show that there is no problem of disappearing agency, and we will see that my response to the first challenge will be conducive to this end. Let me point out, right at the start, that there are many controversial issues concerning the metaphysics of action and causation that are beyond the scope of this chapter. This means that I will have to make some substantial assumptions, especially in the first and second part.

1.1  Part 1: The standard theory

The causal theory of action, as I understand it, consists of two parts: a causal theory of the nature of actions and a causal theory of reason explanation. The former says, basically, that actions are events with a certain causal history (they are actions in virtue of this history). The latter is often stated by reference to Davidson’s claim that reason explanation is a “species of
causal explanation” (1980, p. 3). The core of the causal theory can then be unpacked as follows:

Causal theory of the nature of action (CTA): An agent-involving event is an action just in case it is caused by the right agent-involving mental states and events in the right way.

Causal theory of reason explanation (CTR): Reason explanations of actions are explanations in terms of the agent’s mental states and events that rationalize and causally explain their performance. In particular, a reason explanation of an action in terms of the agent’s mental states and events is true only if those states and events causally explain the action.

The right mental states and events, which CTA refers to, are mental attitudes that rationalize the performance of the action (such as desires, beliefs and intentions). The right way of causation is non-deviant causation. We will return to the issue of non-deviant causation in the third part. For now, consider the following clarifications.

Firstly, it is plausible to think that CTA and CTR stand or fall together, because it is plausible to assume that the mental states and events that CTA refers to are the same mental states and events that feature in reason explanations according to CTR. I will assume, throughout, that CTA and CTR do stand or fall together.¹

Secondly, the conjunction of CTA and CTR is only the skeleton of the causal theory. It is supposed to capture the common core of causal theories of action or, more generally, the core of the causal approach to agency. In the following, I will nevertheless refer to this core as a theory, for the sake of convenience.

Thirdly, the view is usually referred to as the causal theory of action. Strictly speaking, it is an event-causal theory. There are alternative causal theories of agency, and I shall refer to it as the event-causal theory, or the event-causal approach, in order to avoid misunderstanding. As I understand it, an event-causal theory is not committed to the claim that only events are causally efficacious entities. It may allow for the causal efficacy or relevance of states, dispositions and other standing conditions, and it may construe events as particulars or instantiations of properties. But it is committed to the claim that at least one event is among the causes of every effect. Events are thought to play a central role in every causal transaction, because events are the entities that initiate or trigger the occurrence of the effect. Another way of characterizing the kind of causation in question would be to say that it is efficient causation by events, states or property-instantiations (by substances) as opposed to both teleological causation and causation by substances (qua substances).

Fourthly, CTR makes no claims concerning the nature of reasons. In particular, it does not claim that reasons are identical with mental states or
events, and it is compatible with the view that reasons are facts or states of affairs (see Dancy 2000, for instance). CTR says that reason explanations are given in terms of mental states and events that rationalize the action. If reasons are facts, CTR can accommodate this by construing their role in reason explanations as the contents (or as what is being represented by the contents) of rationalizing mental states and events (in true and justified reason explanations). In order to avoid misunderstanding, I will call mental attitudes that rationalize actions reason-states, rather than reasons. Given this, any attempt to reject the event-causal theory by objecting to the claim that reasons are mental attitudes is simply missing its target.

Finally, the theory is sometimes referred to as the desire-belief theory of action. This is misleading insofar as the event-causal approach is by no means committed to the claim that all actions are caused and explained by desires and beliefs. Most proponents of the view hold now that intentions, construed as a genuine type of mental attitude, play a central role in the causation and explanation of action. In principle, proponents of the view may refer to all kinds of mental entities that qualify as an agent’s mental states, mental events, and to relationships between them. My aim here is to defend the event-causal approach to agency in general, rather than a particular version of the view.

1.2 Part 2: Why believe it?

Is the event-causal theory a plausible and intuitively attractive position? It is, I think, neither particularly plausible nor particularly implausible on the grounds of commonsense intuition. The reason for this is that commonsense is silent about most of the issues involved. It seems very plausible to say that the explanation of an action in terms of the agent’s desires, beliefs and intentions is true only if the cited mental states motivated the agent to perform the action. But does this mean that the agent’s mental states must have caused the action? Does this commit us to the claim that actions are events? Commonsense holds no answers, let alone conclusive reasons in favour of a particular causal or non-causal position.

What, then, is to be said in favour of the event-causal theory on the grounds of philosophical argument? Many proponents of the view think that the best argument is provided by Davidson’s challenge from reason explanation. Davidson pointed out that it is possible to rationalize the performance of an action in the light of some of the agent’s reason-states irrespectively of whether or not the agent acted for them. The agent may not have performed the action because of having those reason-states. In other words, reason explanation cannot be reduced to rationalization. Something else is needed.

This point is particularly salient in cases in which an agent has more than one set of reason-states that would rationalize the performance of an action,
Agency, Ownership, and the Standard Theory

but in which there is reason to think that the agent acted only because of one set of reason-states. What explains the fact that the agent acted for one rather than another set of reason-states? In general, and to use Davidson’s expression, what is the “mysterious connection” between reasons and actions? Davidson suggested that this can only be a causal connection. What else could it be? This, in short, is Davidson’s challenge. (Davidson 1980, especially pp. 8–11. Compare also Ginet 2001.)

It has been pointed out that the core of this challenge is a metaphysical rather than an epistemological point (Child 1994, for instance). It is not about how we can know that an agent acted for a reason, but about what makes this true. Causal connections can ground the truth of reason explanations, and they can, in general, explain the metaphysical connections between an agent’s reason-states and actions.

Opponents have responded by pointing out that this challenge provides at best an indirect and negative argument for the event-causal approach. We can distinguish two points here. First, as an inference to the only explanation, the argument lends merely negative support. It supports the view only on the assumption that there are no viable alternative accounts of reason explanation that can meet Davidson’s challenge. Secondly, it is only an indirect argument insofar as it gives no direct and positive argument for the metaphysical framework of the event-causal approach. In particular, it offers no direct support to CTA. It provides negative support to CTR, and gives indirect support to CTA only insofar as CTA and CTR stand or fall together.

Given all this, it seems that one can undermine the force of the argument simply by presenting an alternative theory of reason explanation that meets Davidson’s challenge. Wilson (1989), Sehon (2000), Ginet (1990 and 2001) and Schueler (2003) have pursued this line of argument, and they have offered alternative non-causal accounts of reason explanation. Proponents of the event-causal theory have responded by criticizing the offered non-causal alternatives, and they have argued, convincingly I think, that Davidson’s challenge is very much alive (compare Mele 1992 and Clark 2003, for instance). But I do not think that the case for the event-causal approach rests on Davidson’s challenge alone. In the remainder of this part, I will present a direct argument from the metaphysics of agency.

1.2.1 Naturalism and the event-causal order

Many philosophers, I suspect, would not agree with the suggestion that Davidson’s challenge provides the only argument in support of the event-causal theory. For many, the best reason to endorse the view consists in its apparent compatibility and congeniality with naturalism. According to a fairly widespread form of philosophical naturalism, all particular occurrences, processes, and changes are to be understood and explained in terms of event-causation. In particular, any appeal to substance-causation, irreducible teleology or sui generis acts would constitute a violation of naturalism.
Given a strong commitment to this kind of naturalism, one has, it seems, a strong reason to endorse the event-causal approach to agency, because this view locates or situates agency within the event-causal order, as it were. Opponents of the view will point out, rightly, that this does not give them any reason to endorse the event-causal theory, simply because they reject this kind of naturalism. (This does not commit them to some kind of anti-naturalism. The term naturalism has been used in many different ways. Rejecting the outlined kind of event-causal naturalism, opponents of the event-causal theory need not reject naturalism as such.)

However, if the characterization of naturalism is weakened, in the right way, we can construct an argument for the event-causal theory that does not beg the question. Or so I will argue now. It should, I think, be uncontroversial that human agents are part of the event-causal order in the following weak sense. Our bodily movements are events that are part of the event-causal order in the sense that their occurrence and execution can be explained in terms of event-causation only—in terms of muscle contractions, motor commands, neural activity in the motor cortex and areas of the prefrontal cortex, and so on. This claim is clearly weaker than the claim that all occurrences, including actions, must be explained in terms of event-causation, and I suspect that only very few non-causalists, if any, would object to this weak or minimal version of naturalism. It is undeniable, I think, that we (or our living bodies, if you like) are part of the event-causal order in this weak or minimal sense. I will assume, from now on, that this is the case, and I will express this by saying that human agency is minimally part of the event-causal order. This is the first of two main assumptions for my argument.

Non-causalists often point out that we are primarily interested in actions, rather than bodily movements, when we are interested in human agency. When we give a reason explanation, for instance, we explain the performance of an action, rather than the mere occurrence of a movement. This, they rightly point out, is often overlooked due to the close relationship between bodily movements and basic actions. Roughly, a basic action is something an agent can do without doing anything else.³ It is widely agreed that every action of ours is either a basic action or brought about by the performance of a basic action (perhaps via a chain or tree of non-basic actions). To take a standard example, you can give a signal by raising your arm. If you do so, you perform a non-basic action (giving a signal) by performing a basic action (raising an arm). The basic action is not performed by doing something else. It is, in particular, not performed by performing a bodily movement (the rising of your arm). But the basic action is also not identical with the bodily movement. Not every movement of this type constitutes or realizes the raising of an arm. They may be token-identical, but they are not type-identical.

Proponents of the event-causal theory can agree with all this, and I think they should agree. We are interested primarily in intentional behaviour, not
bodily movement. We perform all non-basic actions by performing basic actions, and basic actions are not type-identical with bodily movements. Nevertheless, it seems obvious that basic actions stand in some intimate relationship with bodily movements. Should they not be token-identical, basic actions are at the very least partly constituted or realized by bodily movements. This appears to be obvious, and I will assume that this is the case. That gives us the second assumption for my argument.

Taken together, the two assumptions give rise to the following central question in the metaphysics of agency: how can human agency be part of the event-causal order at all? Given that overt actions are constituted or realized by movements, and given that bodily movements can be explained in terms of neuro-physiological events, how can agents exercise their agency? How can agents, who are minimally part of the event-causal order, exercise control over their actions?

To make this question clearer, let us briefly consider the case of actions that are done for reasons (in the broad sense of being rationalized and motivated by reason-states). We can distinguish here between four things: actions, an agent’s reason-states, bodily movements, and the causes of bodily movements. On the one hand, these actions are done for reasons. They are performed because their agents have certain reason-states. On the other hand, they are at least partly constituted or realized by bodily movements, which can be explained by reference to neuro-physiological events alone. Given the constitutive relationship between actions and movements, we need an explanation of how the influence of reason-states on the agent’s actions is related to the causal efficacy of the neuro-physiological causes of the agent’s bodily movements.

Agential and rational control is in need of explanation. In the following section, I will introduce what I take to be the basic options in the metaphysics of action, and we will assess them in light of the task that has just been outlined.

1.2.2 The metaphysics of agency

What is an agent? What is agency? First, let me restrict our considerations to overt actions (that is, roughly, actions that involve bodily movement). A good starting point is to think of overt agency in terms of self-movement. Agents are beings or systems that can bring about change in their environment by bringing about change in themselves (by moving in a certain way). Agency is an exercise of this ability.

What is self-movement? Intuitively, it is movement that is brought about or initiated by oneself (by the agent or the system itself). What does this mean? We can distinguish here between three main options in the metaphysics of agency, which give three different answers to this question.

According to the first, self-movement is initiated by the self in the sense that it is caused by salient features of the agent, which are themselves caused
by other features of the agent or the agent’s environment. According to the second, self-movement is initiated, literally, by the agent (by the persisting being that is the agent, rather than features of that being). According to the third option, self-movement is initiated by a mental act of the will, often called volition, which is not itself caused by anything else.

Those, I think, are the three main options in the metaphysics of agency. The first is a reductive approach. It reduces an agent’s role or power to the causal roles or powers of agent-involving states and events. Paradigm instances of this approach are event-causal theories, which provide a metaphysical account of agency in terms of event-causal relations between reason-states and actions.6 The second approach is a non-reductive approach. It construes an agent’s role or power as irreducible. Paradigm instances of this approach are agent-causal theories of action.7 The third approach is also non-reductive in the sense that it rejects the reduction proposed by the first approach. Paradigm instances of this approach are volitional theories of action.8

A fourth possibility is to reject the project of giving a metaphysical account of agency as misguided. On this view, the notion of agency is essentially normative, and it can be captured and understood only from a normative or practical standpoint. The phenomenon of human agency is thought to disappear from any metaphysical, naturalistic or otherwise theoretical point of view. Let us call this the dual standpoint view.9

1.2.3 Dual standpoint theories

Our question is how human agency can be minimally part of the event-causal order. In particular, how can agents, who are minimally part of this order, exercise control over their actions? In order to give an answer to this question, one must provide a metaphysical account of human agency. The dual standpoint view does not acknowledge that there is a need for an explanation, and it rejects the metaphysical quest as misguided. It is, for this very reason, unsatisfactory, as it leaves one of the fundamental questions about human agency unanswered.

This dismissal of the dual standpoint view would be unjustified, if it could be shown that a metaphysics of agency cannot be given. But I have not seen any good argument to this conclusion. On the face of it, a metaphysics of agency is worth wanting, and different proposals are on offer. Given this, I see no reason to adopt the dual standpoint view. (We will return to the charge that agency disappears from a naturalistic standpoint further below.)

1.2.4 Volitionism

Volitionism is widely rejected, mostly due to internal problems with the theory. I will not summarize the most common objections here.10 Let us consider, instead, whether the view can give an answer to our question.