

Quaestiones Infnitae

PUBLICATIONS OF THE DEPARTMENT OF
PHILOSOPHY AND RELIGIOUS STUDIES
UTRECHT UNIVERSITY

VOLUME LXXXVI

Copyright © 2015 by Niels van Miltenburg

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, electronic, mechanical, photocopying, recording, or otherwise, without prior permission in writing of the author.

Cover art: Joyce Overheul

ISBN 978-94-6203-832-5

Printed by CPI-Koninklijke Wöhrmann

Freedom in Action

Handelingsvrijheid

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 9 juni 2015 des ochtends te 10.30 uur

door

Niels van Miltenburg

geboren op 22 juni 1986

te Nieuwegein

Promotoren: Prof.dr. A. VISSER
Prof.dr. T. MÜLLER

The research leading to this doctoral thesis was funded by the European Research Council under the European Community's Seventh Framework Programme. ERC Grant agreement 263227.

Voor mijn ouders

Our remedies oft in ourselves do lie, Which we ascribe to heaven
—Shakespeare*

* *All's Well That Ends Well* (1.1 212–13)

Acknowledgements

As Charles Darwin recognises, writing does not come naturally: “man has an instinctive tendency to speak, as we see in the babble of our young children; whilst no child has an instinctive tendency to brew, bake, or write.”* For me this statement certainly rings true and I feel that I would not have been able to write anything cogent, let alone an entire thesis, without the continuous aid and support of some wonderful human beings.

Thomas Müller, thank you for giving me the opportunity to write this thesis, for meticulously commenting on several versions and for your tireless support. Together with Albert Visser you have given me the freedom to develop my own thoughts, which I consider to be a great privilege.

I have been lucky to write this thesis alongside three extremely intelligent and generous fellow PhD researchers and friends: Antje Rumberg, Dawa Ometto and Jesse Mulder. Dawa, there isn't a single thought in this book that we have not extensively discussed and many new ideas have originated from these discussions. If there is a good idea in this thesis, it is as much yours as it is mine, and I can only hope to have expressed it sufficiently well. Jesse, there is perhaps no one from whom I have learned more about the way in which philosophy should be conducted than from you. Antje, thank you for all our discussions and allowing me to share the ups and downs of writing a thesis with you.

I am grateful to Florian Fischer for giving me a crash course in the metaphysics of powers and for our subsequent joint work. Daan Evers, thanks for our many conversations about free will and neuroscience, and your open mind in discussing all things philosophy.

Many others have had an impact on this thesis by offering me written or spoken feedback. My gratitude goes out to, among others, Charlotte Alderwick, Marius Backmann, Michael De, Kim Frost, John Hyman, Annemarie Kalis, Erasmus Mayr, Timothy O'Connor, Sarah Paul, Herman Philipse, Helen Steward, John

*The *Descent of Man*, 1871, Chapter 3.

Schwenkler, Markus Schrenk, Verena Wagner, René van Woudenberg and Leszek Wroński.

I want to thank everybody at the department of Philosophy and Religious Studies in Utrecht, both for my philosophical education and for providing an excellent research climate. A special thanks goes out to the current group of research master students, Janneke van Lith, Biene Meijerman and Suzanne van Vliet.

Thanks to all my friends, musical, gaming or otherwise, for providing the much needed distractions and making my live a joy.

I am grateful to my parents, Gerard and Marjo, and my siblings, Mark en Karin. You have raised me to be an independent thinker and have always supported my endeavours.

Above all, I want to thank Belén, you have been with me from the moment I first started studying philosophy, you have always believed in me and kept my feet on the ground. There are no words to express your importance to me.

Niels van Miltenburg
April 2015

Contents

Introduction: A Simple Picture of Free Will	1
Method	3
Aim and outline	12
1 Free Will in the Age of Neuroscience	19
1.1 Libet's experiment	21
1.2 Technical issues	24
1.3 Some conceptual issues	27
1.3.1 Initiation and determination	28
1.3.2 Generalisation	31
1.3.3 Alternative interpretations	36
1.4 Haynes's experiment	39
1.5 Conceptual worries about Haynes	44
1.5.1 Haynes and alternative interpretations	45
1.5.2 Haynes and determination	46
1.5.3 Haynes and generalisation	50
1.6 The futility of Libet-style experiments	52
1.7 Priming and illusionary willing	54
1.8 Concluding remarks	58
2 Free will and Control	63
2.1 The free will problem	65
2.2 CTA and the problems of control	72
2.3 Free will as intentionality+ X	79
2.4 Free action as intentional action	86
2.5 Concluding remarks	88
3 Accidentality in Action I	91
3.1 The causal theory of action	94
3.2 The disappearing agent	98
3.3 Deviant causal chains	102
3.3.1 Causation in the right way	104
3.3.2 Representation in the right way	107
3.3.3 Causation because of representation	110
3.3.4 Miscellaneous strategies	115

3.4	A diagnosis	123
3.5	Concluding remarks	130
4	Accidentality in Action II	133
4.1	Non-causal action theory	134
4.1.1	Reasons as context placing explanations . .	134
4.1.2	Teleological realism	140
4.1.3	Intrinsically intentional decisions	144
4.1.4	Concurrent intentions	145
4.2	Agent-causal accounts	147
4.2.1	Steward's agency incompatibilism	151
4.2.2	Clarke's integrated agent-causal account . .	155
4.2.3	O'Connor's agent-causal powers	157
4.2.4	Lowe's volitionalism	162
4.3	Concluding remarks	166
5	Practical Knowledge	169
5.1	The question 'Why?'	171
5.2	The cause of what it understands	176
5.3	Practical knowledge and side-effects	181
5.4	The problem of failed action	187
5.5	The mistake is in the performance	193
5.6	A constitutive account	197
5.7	Concluding remarks	207
6	Powers, Processes and Practical Knowledge	211
6.1	Contemporary metaphysics of powers	213
6.2	Reduction and prevention	216
6.3	Realism and prevention	220
6.4	Manifestations as processes	225
6.5	The form of processes	233
6.6	Powers, indeterminism and control	242
6.7	Practical knowledge and determinism?	247
6.8	Concluding remarks	251
	Conclusion	253
	References	259
	Samenvatting in het Nederlands	289

Introduction: A Simple Picture of Free Will

One of the many reasons, I believe, why philosophy falls short of a satisfying solution to the problem of freedom of the will is that we still cannot refer to an unflawed statement of libertarianism.

—David Wiggins¹

We are free. We have the capacity to select one course of action from among alternatives and execute it. Philosophers call this capacity of choice ‘free will’. At this point agreement among philosophers ends. For over two millennia a plethora of questions concerning free will have been heavily debated. The most notorious of these, is the question whether this capacity is threatened by determinism. Determinism is, roughly speaking, the idea that there is only one possible future: what will happen is just as fixed as what has happened.² It is easy to see why someone would think that determinism excludes free will. After all, if there is only one possible future, how can there be a capacity to select one among

¹Wiggins [1973, pp. 269–270].

²In the history of philosophy, determinism has been defended on many different grounds, the most prominent of which include: the idea that the future must be determined because God already knows what will happen, and the idea that every future event is necessitated by past events in conjunction with the laws of nature.

many? Philosophers who believe that we have free will and that the doctrine of determinism must therefore be rejected are called *libertarians*. Their main adversaries are those who believe that free will and determinism can co-exist; they are aptly referred to as *compatibilists*. The clash between libertarians and compatibilists is one of the fiercest and longest in philosophy. Currently the majority of professional philosophers sympathises with compatibilism.³ An interesting fact, since to laymen, libertarianism seems to be the more intuitive position.⁴ Finally there is a growing army, led by contemporary neuroscientists, of those who deny the existence of free will altogether.⁵

I aim to defend a very straightforward and simple picture of free will: whenever we act, we actualise one of the real possibilities open to us. The universe is, as it were, like Borges's garden of forking paths, and human beings have the power to decide, at each juncture, what road they will follow. In this way we determine what will happen and hence the future of the universe is shaped, among other things, by our choices.

This simple picture construes the alternatives from which we are able to select as *really* possible, not as merely conceivable or only theoretically possible. It follows that the simple picture is a *libertarian* portrayal of free will. For if there is more than one real possibility—more than one thing that can really happen—the future must be open. Despite my libertarian sympathies, I believe that many of the contemporary versions of libertarianism defended in the current debate obfuscate rather than elucidate the simple picture. But before we get into the thick of the philosophical

³See Bourget and Chalmers [2014].

⁴At least in so far as experimental philosophy is to be believed: “Our hypothesis is that people have an incompatibilist theory of moral responsibility that is elicited in some contexts but that they also have psychological mechanisms that can lead them to arrive at compatibilist judgments in other contexts” [Nichols and Knobe 2007, p. 664].

⁵A more fine-grained categorisation of the existing positions in the debate will be provided in Chapter 2.

battle, I would like to take a step back and reflect on the kind of enterprise an inquiry into the concept of free will is, or should be.

Method

The investigation into free will conducted here, is first and foremost a philosophical endeavour. This is not the place to offer, let alone argue for, any theory on the nature of philosophy. However, I do want to repeat the words Elizabeth Anscombe once used to characterise the subject in a university prospectus: “Philosophy”, she said, “is thinking about the most difficult and ultimate questions”. What is striking about this—as her daughter Mary Geach observes—is that Anscombe did not see the need to further qualify the word ‘thinking’ by adding adverbs like ‘systematic’, ‘analytical’ or ‘critical’.⁶ Philosophy is *just* thinking. Since thought can take many forms, philosophy does not strictly prescribe the use of any single methodology. It instead allows for a wide variety of methods: from conceptual analysis to dialectics, and from the use of thought experiments to comparative analysis. All of these methods will be employed in the current investigation: the only condition on the kinds of thinking invoked is that they will, I hope, get us closer understanding free will.

Secondly, I would like this investigation to be a *metaphysical* inquiry. Aristotle characterised metaphysics as “the study of being *qua* being” [Aristotle 1998, pp. 79-80, 1003a21-31]. But what does this mean? Aristotle’s wording could suggest that there is a special subject matter studied by the metaphysician: “being *qua* being”, i.e., ‘being itself’ or *das Sein*. However, we can also understand Aristotle’s remark as involving three instead of two things: a study, a subject matter (being) and a manner in which the subject matter is studied (*qua* being).⁷ Take a cat, for example. It can be

⁶See Geach’s preface to [Geach and Gormally 2005].

⁷This is the interpretation of e.g., Cohen [2014].

studied as a material object in physics, as an organism in biology, or as an entity of modern-day civilisation by sociology. Metaphysics goes beyond these sciences and studies it as a being, just in so far as it exists.⁸ In other words, when we metaphysically consider something we want to know what it *really is*. Hence, Barry Stroud's somewhat less cryptic description of metaphysics as "the quest for reality" [Stroud 2000]. I would like to find out what free will really is and therefore I say that I want this investigation to be metaphysical.

But is it not the case that every philosopher who occupies herself with the study of free will wants to know what it really is? What, then, sets a *metaphysical* investigation into free will apart from other forms of enquiry? To elucidate these matters, it is helpful to sketch a non-metaphysical approach to free will and point out what I find objectionable about it. The approach I have in mind is exemplified by Peter Strawson in his seminal paper 'Freedom and Resentment' [Strawson 1962]. Although the word 'freedom' figures prominently in the title of his text, Strawson mainly concerns himself with moral responsibility and whether or not our practices of praising and blaming each other for the results of our actions could be justified if determinism was true. Is it rational to praise someone for something she would have done anyway? If there was but one possible future and there would thus be nothing else she could have done, then how commendable is it that she did just that thing? Strawson expresses sympathy both for those who believe that praising and blaming is compatible with determinism and those who think it is not. But when he asks himself to which of the parties he ultimately belongs, he states that he first and foremost belongs "to the party of those who do not know what the thesis of determinism is" [ibid, p. 1]. However, this does not

⁸And what it is to be, might of course be different for a cat than for a planet, a tablespoon, a rock, a sentence, a human being or the object of study in this thesis: a capacity.

stop him from criticising libertarians for overintellectualising the issue of moral responsibility.⁹ What he means is that incompatibilists seem to think that one can only be justified in holding somebody morally responsible if the person satisfies some objective requirements of being morally responsible, that is, if he *really is* morally responsible. But our practices of praising and blaming, Strawson argues, do not need such metaphysical grounding. Instead, they are grounded in the *reactive attitudes* we have towards each other's behaviours: immediate reactions like gratitude, anger or resentment. According to Strawson, these attitudes are natural and irrevocable. They are part and parcel of our interpersonal way of life.

But would it not be rational to try to change this way of life, if we found out that determinism was true?

[O]ne who presses this question has wholly failed to grasp [...] the nature of the human commitment that is here involved: it is useless to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do. [Strawson 1962, p. 19]

Therefore, Strawson claims that the practice of holding each other responsible “neither calls, for nor permits, an external ‘rational’ justification” [ibid, p. 25]. And hence, he is able to conclude that there can be no conflict between moral responsibility and determinism, without engaging into a metaphysical study of these notions.

The first thing to note about Strawson's approach is that his main concern is with the vindication of our practices. But is there not more to philosophy? Recall Anscombe's characterisation of philosophy as thinking about the most difficult and ultimate ques-

⁹He also criticises a specific type of consequentialist compatibilists. Thus he is arguing in much of the paper that both sides of the debate are mistaken. Nevertheless Strawson himself ends up defending what can clearly be called a compatibilist position. For the sake of both brevity and simplicity, I will not go into these intricacies right here.

tions. Metaphysical questions are nothing if not difficult and ultimate. A philosopher, it seems to me, cannot help but think about these questions. Even if all our practices were validated, she would not rest happily. She would still like to know what, for instance, our capacity of choice *really is*. But perhaps Strawson is sceptical that we can, just by thinking, find out much about what things fundamentally are. At least his own philosophy, which he calls ‘descriptive metaphysics’, is focused on investigating the way we think and talk about reality, rather than directly concerned with reality itself.¹⁰ Although descriptive metaphysics might be an interesting sort of enquiry of its own, we should note that our thinking does not proceed in a vacuum. We are not just thinkers, after all, but also observers and agents, and our thoughts are full of concepts we grasp through our observations of, and our dealings with reality. In thinking we directly engage with the world.¹¹

The second and more important issue, is that it is impossible to radically disengage practice from metaphysics. When I find out that I was hit in the face by someone, not out of malevolence, but because that person has an uncontrollable tic, my initial anger subsides and is replaced with an attitude of understanding. The next time that person hits me, my initial reaction might not even be anger, I might have an immediate attitude of understanding, or I might feel pity. So why can it not be the case that I would also react differently if I knew that all the behaviours of the people around me were physically determined? Perhaps Strawson would argue that, although it is possible that our reactive attitudes in certain specific cases are subject to revision, it is impossible to change *all* our initial reactions, since they are indispensable to our

¹⁰See Strawson [1959].

¹¹If it is correct that the structure of reality is mirrored in the structure of our thought, then even descriptive metaphysics might yield insight in the way things really are, though it would then cease to be the deflationary kind of enterprise Strawson takes it to be. For an excellent defence of such conceptual realist thoughts, see Mulder [2014].

way of life. Such claims about what is natural or indispensable to the lives humans lead might well be right—although they reek somewhat metaphysical. But why conclude from the impossibility of radical change in our attitudes that metaphysical considerations are separate from considerations of practice? Suppose that we find out that the practices Strawson describes are indeed indispensable to our form of life, and additionally discover that we can only make sense of the existence of these practices and such a way of life against the background of indeterminism, then it seems that we should not conclude that the question of (in)determinism is irrelevant. To the contrary, we should conclude that the world must be indeterministic: how else could it be possible for us to lead the lives we manifestly lead?¹²

Even though most philosophers do not disengage practice and metaphysics as radically as Strawson does, many still start the analysis of free will on the level of practice. As Helen Steward [2012a] points out, it is common nowadays to start an investigation into free will by noting that the term ‘freedom’ has many meanings and subsequently attempt to zero in on the conception of freedom relevant to free will by asking what kinds of freedom might be important for moral responsibility, or which kinds are worth wanting. It is easy to see the appeal of these strategies. As Steward writes:

[O]ne does not want to debate the question whether or not something called ‘freedom’ is compatible with determinism in a complete void, [and these strategies] might seem sensible ways of making an unhelpfully indeterminate notion more determinate [ibid, p. 5].

¹²John Hyman once expressed (in personal conversation) that his main issue with libertarianism is that libertarians believe that they can tell something about the fundamental nature of the universe (i.e. whether it is deterministic or not) on the basis of mere reflection on the capacities of human agents. But what would be so striking about that? Are human agents not also part of the universe?

But the danger of this approach is that we could end up getting no further than discussing the particularities of rather exalted freedoms like moral autonomy, or freedom of speech. We might never get down to the fundamental level and never find out what free will really is. Moreover, we might be unable to correctly answer questions about these higher freedoms, without first understanding the metaphysically fundamental notion.

In fact the ‘starting from practice’ approach often complicates matters more than it elucidates them. Consider the starting observation made by many philosophers that you cannot be free when you are forced to do something. Prisoners have their freedom taken away by being forced to stay in a correctional facility. But should we say that they lack the capacity of choice? Certainly the alternatives they can choose from do not include leaving the facility: their alternatives are limited. But in this regard prisoners are not different from someone who has fewer options because she is poor, or missing a leg. Nevertheless it seems that even prisoners are still free to choose from among the limited alternatives they do have. But could there not be cases where an agent only has one alternative? What if someone points a gun at you and threatens to shoot unless you give her your ice cream? In this situation giving up the cone clearly seems the best action, but could you really not resist and try to overpower your attacker? And what about a situation in which you are forced at gun-point to press the button that launches all the world’s nuclear missiles? Answering such questions seems really difficult to me, especially without first having a better grasp of what the capacity to select alternatives really is. It seems plausible that taking away alternatives could impinge on our autonomy, but does it undermine our fundamental capacity of choice?

Another line often taken in the philosophy of free will starts from the observation that agents who are truly free must be able

to express themselves. It follows that an action is only free if it is an expression of the agent's personality. Indeed it is commendable to allow or even stimulate human beings to express their own character—artistic freedom, for instance, should be well preserved in every society. But is this a proper starting point for an enquiry into the capacity of free will? If I decide to actualise an alternative that was out of character for me, didn't I do so freely? Are we *not* morally responsible for killing someone, if the act of murder was out of character? This, of course, is not the place to suggest that philosophers could not be perfectly able to come up with answers to these questions, or with answers to the above questions on coercion and the limitation of alternatives. However, the considerations involved in thinking about these questions might well be very different from considerations relevant to understanding the fundamental capacity of choice. It makes sense, for instance, to say that someone is more free when she is better able to express herself, or that she is freer when more alternatives are open to her, but does such a gradable notion of freedom make sense when we discuss the very capacity to select from alternatives? On that level, it seems that you either recognise multiple available choices and are able to select one of them, or that you are not so able.

Although the question of what free will really is has not disappeared from the debate, it is remarkable that philosophers feel the need to settle their differences on the basis of intuitions concerning, for instance, moral responsibility.¹³ The commonality of the 'starting from practice'-approach to freedom, perhaps makes sense from a compatibilist perspective. After all, compatibilists might think that there is no real choice on the fundamental level, because on that level everything is in fact determined (or might as well be). Despite such determination we can still be free, they

¹³The enormous amount of literature on whether Frankfurt's [1969] cases do, or do not rebut libertarianism is perhaps the most vivid illustration of this tendency.

argue, because not all physically determined actions need to be coerced or out of character or something such. For compatibilists, then, philosophy cannot do much more than explicate how to make sense of the more exalted conceptions of freedom, simply because they think that there is no such thing as freedom at the fundamental level—the concept of freedom does not apply there. Still, even libertarians who deny that freedom is compatible with everything’s being fundamentally determined, often start from a higher level. Robert Kane, who has been one of the main torch bearers of libertarianism in the last few decades, for instance, claims that real freedom occurs only when we make important life-altering decisions. The rest of our actions merely derive their freedom from these “character forming actions”.¹⁴

All in all, it is striking how the contemporary free will debate rages on without giving much attention to the fundamental nature of action, or of (in)determinism. The philosophy of action is actually seen as a different branch of philosophy altogether, and questions concerning determinism are as often as not left for the physicist to answer. The free will debate seems to proceed on the assumption that everyone agrees on what is meant by ‘action’ or ‘indeterminism’, or that we at least have some intuitive grasp of these concepts.¹⁵ This thesis will try to be different. In fact for a thesis on free will, there will be very little reflection on freedom. The reason for this is that the simple picture of free will only requires that an agent has the power to control which option she will actualise. And that power, I will argue, is nothing more than the capacity to act intentionally. Therefore, there is no room for a notion of *free* action separate from the notion of intentional action as such.¹⁶ Understanding free will comes down to understanding

¹⁴See Kane [1985, 1996, 1999b].

¹⁵Helen Steward’s recent book [2012a] is a notable exception to this rule.

¹⁶Unless, of course, when ‘free’ is intended to refer to autonomy or any other more exalted notion of freedom. What I mean is that there is no such room for a *metaphysical* notion of free action.

agency, and hence, the free will debate should coincide with the philosophy of action.

But before we turn our attention fully to the subject matter of this thesis, there are some final remarks to be made on metaphysics and how it should be conducted. Metaphysics often involves generalisation. We abstract away from individual choices to get to know something about the capacity of choice in general, and we may abstract even further to understand the nature of capacities in general. But it is important that we do not confuse aiming for generality with aiming for parsimony—the kind of feature one would sometimes like in a logical system. What if reality is fundamentally not like a desert, but more like a jungle? What if it is inhabited by all kinds of complex entities, instead of by simple objects? Of course, Ockham’s razor can be a useful tool for selecting one theory from amongst two theories with more or less equal explanatory power, but we should be careful not to use it to shave off entities for the sake of parsimony but at the cost of understanding.¹⁷ Metaphysics, like the rest of philosophy, comes down to ‘just thinking’. And as remarked above, thinking does not proceed in a vacuum. Therefore we might sometimes be unable to abstract further without losing sight of the thing we were originally thinking about.

Earlier in this section we saw that metaphysics is concerned with questions of how things really are. But when philosophers, after long and arduous work, come up with an answer to a specific metaphysical question, we are often left wondering how it is possible that the thing in question really is as the philosopher conceives of it. As Robert Nozick [1981, p. 8] points out, “many philosophical problems are ones of understanding how something is or can be possible”. These ‘how possible?’ questions, in turn, often force one to try to explain how other things really are. The

¹⁷Cf. John Heil [2012, p. 79]: “Parsimony figures in the endgame, not at the outset of theorizing.”

current investigation into free will, will not be different. I will not just try to come up with the proper understanding of the simple picture. I will also try to show how it fits into our understanding of the world as a whole. However, this endeavour in itself is again bound to raise more questions than it answers. And that is what makes philosophy so hard—solving one matter immediately leads to questions on another matter. But this difficulty, I think, should not stop us from wanting to know what the world fundamentally is like. It should not prevent us from embarking on the quest for reality.

Aim and outline

In the beginning of this introduction, I sketched a simple and straightforward picture of free will: when we perform a (free) action, we actualise one of the real possibilities open to us. Many philosophers would find this picture of free will overly naive. They would argue that we need to adjust, or even altogether replace, this picture. Compatibilists, for instance, would say that the possibilities we see before us when we act, do not necessarily need to be really open to us. According to them, we could still be free if only one future was really possible. Some of them contend that it is sufficient for free will, that we believe or think that multiple possibilities are open to us. Freedom is only in danger, they claim, when we feel that some of the possibilities are artificially taken away from us, by coercion for instance. But even libertarians propose alterations to the simple picture. So called ‘deliberative libertarians’, for instance, believe that actions are only free when they are preceded by conscious deliberation in which an evaluative judgement is made or a preference is formed.¹⁸ As we saw above, other libertarians, like Kane, argue that (directly) free actions can

¹⁸The label ‘deliberative libertarianism’ is from Clarke [2003], he cites Mele [1995] and Ekstrom [2000] as exemplifying such a view.

only be performed in cases of fundamental conflict between desires about what kind of person one wants to be. Hence only a subset of the actions an agent performs are directly free, the rest is only derivatively free—if they are free at all.

This thesis takes a different stance towards the simple picture of free will. I will not be concerned with replacing or adjusting the picture. Instead, I will try to understand and defend the simple picture as it stands. Such a defence will not only involve the rebuttal of both scientifically as well as philosophically inspired criticisms of the simple picture. It will also involve explaining what agency really is like and how the world must be in order for such a picture to be true. The result of this will hopefully be a “reasonable libertarianism” [Wiggins 1973]. Such a libertarianism should not just be wanted for its own sake, but also because:

[c]ompatibilist resolutions to the problem of freedom will always wear an appearance of superficiality, however serious the reflections from which they arise, until what they offer by way of freedom can be compared with something else, whether actual or possible or only seemingly imaginable, that is known to be the best that any indeterminist or libertarian could describe. [ibid, p. 270]

Of course I do not claim that my offering in this thesis will be the best any libertarian could do. However, I do hope that my work can contribute a little to the search for a proper libertarian account.

This thesis consists of six chapters. The first chapter clears the way for a philosophical investigation into free will. I will be arguing against those who believe that advances in contemporary neuroscience show that free will, whatever picture one might sketch of it, is no more than a philosopher’s fantasy. The chapter is not concerned with defending the simple picture view of free will against the challenge from neuroscience—after all, I have not yet

detailed the view at this point. Instead this chapter has a wider scope, in that it tries to show that neuroscientific considerations cannot do much to disprove *any* of the contemporary accounts of free will. The first part of this chapter gives an overview of the many reactions that have arisen to Benjamin Libet's famous experiments [Libet *et al.* 1983, Libet 1985]. These experiments are often taken to show that we lack freedom of the will, because our brains make our decisions before we become aware of them. Although the reactions to Libet's experiments by philosophers show that it is implausible that these experiments really prove that we lack free will, many contemporary neuroscientists stress that their field has advanced a lot since the early eighties and that philosophers are sluggishly lagging behind by still discussing Libet's results. Especially the recent work of John Dylan Haynes's group [Soon *et al.* 2008] is taken to radicalise and strengthen the challenge to free will. I will take up the glove discussing this experiment and argue that, surprisingly, many of the criticisms levelled against Libet can be carried over to Haynes's experiment. I will also attempt to provide a novel argument that shows how no experiment of the sort that Libet and Haynes have conducted can ever successfully disprove the existence of free will.

In the second chapter I will reflect on how the contemporary participants in the free will debate understand the *control* an agent has over her own free actions. We will see that almost all contemporary philosophers conceive of free action as intentional action plus something extra.¹⁹ Interestingly the debate between contemporary libertarians and compatibilists is almost entirely focused on what the 'something extra' is that turns a merely intentional action into a free action and on whether the extra is compatible

¹⁹Putative candidates for this extra requirement are: that the agent must be reasons responsive [Fischer and Ravizza 2000], that the actions must be self-forming, or be derived from self-forming actions [Kane 1999b], or the favourite libertarian requirement: that the agent must have had alternative possibilities of acting.

or incompatible with determinism. Neither side offers much reflection on the nature of intentional action; in fact the majority of both camps seems to agree that some sort of event-causal story like the one offered by Davidson [1963] must be true. But that event-causal theory of action is, I will argue, inherently compatibilist, and for that reason libertarians are bound to lose the contemporary debate by default. Therefore, this chapter will conclude, the advocate of the simple picture should seek to undermine the event-causal account of agency and offer an alternative theory of action.

In the third chapter I take up the task of undermining the event-causal theory of action. The argument of this chapter is that the so-called *problem of deviant causal chains*, which is a well known challenge to the causal theory, is not an isolated problem to be dealt with by suitably refining the causal theory. In fact it is only a symptom of a deeper flaw that consists in the radical distinction causal action theorists make between the *occurrence* of an intentional action and its *rationality*. The former is fully accounted for in causal terms, whereas the latter is wholly explained in terms of representation. I will argue that this separation of causality and rationality, blocks any way of specifying what would be a right, or non-deviant causal chain. Even when causation and representation coincide, this may only be by accident. Furthermore, the causal action theorist cannot give up her radical distinction without at the same time abandoning the project of causal action theory altogether. Therefore, the problem of causal deviance turns out to be an inescapable problem. To add some force to this claim, I will try to minutely show how all of the attempts to specify ‘causation in the right way’ have failed.

The fourth chapter starts the search for an alternative theory about what it is for an agent to act. Two kinds of alternatives are proposed in the literature: non-causal theory and the agent-causal

theory. But neither theory, I argue, delivers a satisfactory account of agential control. I will start with discussing multiple non-causal theories and argue that they all fail for the same reason: they say nothing whatsoever about the material occurrence of the bodily movements that are our actions. Contemporary agent-causal accounts show more promise on this score. Their proponents argue that (free) actions are directly caused by the agent herself. Where agent-causation used to be thought of as an altogether incomprehensible concept, recent agent-causal accounts have made great strides to demystify the notion. Contemporary agent-causalists now argue that agent-causation, or better substance-causation, is not restricted to the realm of the personal or the animate, but also plays a role in the sphere of non-living causal processes. However, defending the view that agent-causation is such a ubiquitous phenomenon comes at a cost: contemporary agent-causalists have a hard time accounting for the specifically agential control exhibited in intentional action, given that their preferred notion of causation appears also in inanimate nature. Most agent-causalists try to overcome this worry by offering a supplemental story about how an agent can cause her actions for a reason. But these additional stories fail to satisfy, I will argue, precisely because they are supplemental, i.e., because they account for the rationality of an action in separation of the causal story about the action's occurrence. Hence, this chapter will conclude that contemporary agent-causal theories suffer from the same separation between rationalisation and causation as their event-causal counterparts.

Is there any account that does not separate causality and rationality? In the fifth chapter I will argue that such an account is presented by Elizabeth Anscombe in her book *Intention*. She shows, I believe, how one and the same happening can be both physical as well as intentional— “a movement that *is* a thought”.²⁰

²⁰This phrase is from Rödl [2007, p. 18].

Understanding *Intention*, however, is no trivial task. As I understand her, Anscombe argues that intentional action is characterised by a very special sort of self-conscious thought of the agent: *practical knowledge*. Moreover, such practical knowledge determines what happens by being *constitutive* of the actions we perform, it is not just productive of our actions it also constitutes their *unity*. In this chapter I will defend the, for now admittedly very cryptic, idea that intentional action is constituted by practical knowledge against contemporary criticism. And conclude that Anscombe's theory finally delivers a satisfactory account of agential control, that can help us understand the simple picture of free will.

The sixth and final chapter is concerned with 'how possible?'-questions. How is practical knowledge possible? And how can it be constitutive of our actions? I will suggest that we can best understand the capacity for practical knowledge as a *power*. Unfortunately, contemporary metaphysicians have so far failed to deliver a satisfactory realist account of powers. This, I will argue, is due to the fact that most philosophers think of the world as a succession of events and of powers as the capacity to produce such events. I propose that to properly understand powers we should accept the category of *processes* as a fundamental ontological category, and think of powers as the capacity to produce such processes. The resultant metaphysics can nicely accommodate for practical knowledge as the power to produce actions. Furthermore, the ontology of powers and processes yields a very natural understanding of indeterminism: the real possibilities of the open future are grounded in the powers of objects. A possibility is actualised when a power is exercised. Hence, when we manifest our power to act we actualise one of the real possibilities open to us. That is freedom in action.

Chapter 1

Free Will in the Age of Neuroscience

You say: I am not free. But I have lifted my hand and let it fall. Everyone understands that this illogical reply is an irrefutable demonstration of freedom.

—Leo Tolstoy²¹

Participants in Benjamin Libet's famous experiments [1983, 1985] were asked to spontaneously perform a simple motor action, such as flexing their wrist or fingers. The test subjects were free to decide when they would perform the action, but had to report at what moment they made the decision. While the subjects performed their actions Libet used electroencephalography (EEG) to measure electrical activity in regions of the brain associated with motoric action. His results were astonishing. He was able to detect a so-called readiness potential about 350 milliseconds before the agents reported that they made the decision to move. Could it be that when the subjects reported that they decided to move they

²¹ *War and Peace*, 2nd epilogue, Chapter 8.

were merely becoming aware of a decision their brain had already made for them? Could this be the case for all human actions, so that we should conclude that we never really decide what to do and when to do it? Could this mean that we lack free will?

Many neuroscientists and some philosophers indeed believe that the results of Libet’s and similar subsequent experiments show that free will is nothing but an illusion. Indeed, as Bill Klemm notices, the idea that free will does not exist “seems to be acquiring the status of dogma” [Klemm 2010, p. 47]. Although the scientific articles that publish the supposedly destructive results are written cautiously and tend to be restrictive about their scope, some neuroscientists have lately been quite vocal outside the strictly academic realm in expressing the opinion that free will is illusory. In the Netherlands alone books like *Wij zijn ons brein* (We are our brains) [Swaab 2010] and *De vrije wil bestaat niet* (Free will does not exist) [Lamme 2011a] are immensely popular.²²

Of course, many philosophers—but unfortunately not many fellow neuroscientists—have responded to the neuro-attack on free will. These authors have mainly pointed out that the scope of the actual neuroscientific experiments is indeed very limited and do not warrant scepticism about free will. However, neuroscientists protest that these advocates of free will focus too much on Libet’s experiments and that their criticism falls short of newer and even more shocking results obtained by fMRI experiments.²³ Furthermore, they argue, these results are only part of a growing body of scientific evidence against free will, which includes experiments on ‘psychological priming’ and research on brain damage.

In this chapter I will try to meet both complaints. I will describe the more recent fMRI experiments (§1.4) and provide a de-

²²And these books are part of an international tsunami of similar titles, e.g. Wegner [2002], Roth [2003], Harris [2012], Gazzaniga [2012].

²³A method that measures activation of brain areas by detecting blood flow.

tailed analysis which attempts to show that, unlike what popular neuroscience suggests, many of the criticisms of Libet's original experiment do carry over (§1.5). I will also go one step further and argue that disproving free will by means of performing any Libet-style experiments is in fact impossible (§1.6). Finally I will attempt to show how the other 'evidence' provided by the free will sceptic can be swept aside (§1.7). But before I can go into these issues, I will explain Libet's findings in section 1.1 and rehearse the most important criticisms thereof in section 1.2 and section 1.3. One remarkable facet of the arguments against free will that neuroscientists present, is that they hardly ever spend any time explaining what they understand by 'free will'. In fact, there is very little reflection in general on many concepts the neuroscientists easily employ, like: 'intention', 'decision', 'urge', 'consciousness' and 'action'. Although this lack of conceptual rigour sometimes leads to confusion which forces one to clarify the meaning of said concepts, I will in general try to stay close to the debate and suppose that we have a more or less intuitive grasp of these concepts.

1.1 Libet's experiment

In 1983 Benjamin Libet and his colleagues published a paper on the timing of the intention to act in relation to the onset of brain-activity. In this paper they wanted to expand on the work of Kornhuber and Deecke [1965] who had discovered that it was possible to use EEG to detect a small negative potential shift along the scalp (a so-called readiness potential) about a second before a self-initiated simple motoric action. In other words, it is possible to detect specific electrical activity in the prefrontal cortex before somebody voluntarily, say, moves a finger. Libet *et al.* wanted to know whether the appearance of preparatory brain processes at such relatively long times before the action indicates that an agent's "awareness of the voluntary urge or intention to act also

appears with such similar advance timings” [Libet *et al.* 1983, p. 624]. To test this, they asked subjects to perform a quick, abrupt flexion of the fingers and/or the wrist of their right hand, while monitoring the time at which they became aware of the voluntary urge or intention to do so. In order to monitor that time, subjects had to look at the screen of a cathode ray oscilloscope (CRO) during the experiment. The CRO was used as a clock, on its screen was a revolving dot that completed a full revolution every 2.56 seconds. After each run of the experiment subjects had to recall the ‘clock time’ at which they became aware of their intention. During the experiment the subjects’ brain activity was measured using EEG and the onset of muscle activity in the wrist was detected by using electromyography (EMG). In this way the time of awareness or time W , as Libet *et al.* called it, could be compared to the time of the readiness potential (RP) and the time of the actual movement. After testing five different subjects in several sessions each consisting of approximately 40 trials, it turned out that on average, conscious awareness came about 350 ms after the onset of readiness potential (and about 200 ms before the act).²⁴ Do these results pose a threat to free will? Libet himself did not think they did. He believed there might be a role for free will in vetoing already initiated actions. There is room for a ‘free won’t’, so to speak. Nevertheless, as far as action initiation is concerned Libet believes that his results contradict the naive and widely held view that we consciously and freely initiate our own movements.

[For if] a conscious intention or decision to act [would] actually [initiate] a voluntary event, then the subjective experience of this intention should precede or at least coincide with the onset

²⁴It is hard to pin down an exact average since Libet *et al.* distinguish between multiple different types of readiness potential and awareness moments. For a full discussion of the results, see the original article [Libet *et al.* 1983] and the subsequent [Libet 1985].

of the specific cerebral processes that mediate the act. [Libet 1985, p. 529]

Thus, insofar as spontaneous and quickly performed motor action is concerned, we ourselves do not decide when to act, but instead “the brain “decides” to initiate or, at least, to prepare to initiate the act before there is any reportable subjective awareness that such a decision has taken place” [ibid, p. 536].

To me it is pretty unclear what it means that the brain instead of the agent “decides” something.²⁵ Especially if we believe the neuroscientists that herald the view that we *are* our brains.²⁶ But to be fair, it is just as unclear what it means that ‘we ourselves’ consciously control or initiate our actions. To understand better the potential threat Libet’s results pose to free will, it might be helpful to consider the two main contenders in the current free will debate: libertarians and compatibilists.

Libertarians believe that free will requires that what an agent does, is not determined before she decides to do so. Her decision can only be a genuine decision if there are genuinely open options from which she can choose. Hence, if Libet’s results show that our brain already determines what we are going to do before we ourselves can make a decision, then it seems that there is no room for free choice. Compatibilists, on the other hand, believe that an action can be free even when its occurrence is already determined by prior events. As long as the causal route up to the action somehow goes through the agent’s conscious decisions, and this might very well be the case since the awareness of the decision precedes the action by 200ms, the agent could still be in control of her own actions, and thus still be free. Therefore Libet’s results should not

²⁵Bennet and Hacker [2003] convincingly show that it is nonsensical to apply verbs like ‘deciding’, ‘learning’ and ‘being aroused’ to the brain, since only agents can learn, decide or be aroused. They also argue that such brain talk is not simply innocent *façon de parler* but leads to conceptual and even empirical errors.

²⁶See [Swaab 2010] for a popular defence of this.

directly seem worrisome to the compatibilist.²⁷ However, some have claimed that Libet’s experiments show that intentions are mere epiphenomena that are caused, but do not themselves cause anything further. If that is true, Libet’s results should be more unsettling to compatibilists who believe that agential control can be spelled out in terms of causation by reason-states:

The position of conscious free will in the time line suggests perhaps that the experience of will is a link in a causal chain leading to action, but in fact it might not even be that. It might just be a loose end—one of those things, like the action, that is caused by prior brain and mental events. [Wegner 2002, p. 55]

Luckily, as I will show in much of the remainder of this chapter, the results of Libet-style experiments do not in the least bit warrant the conclusion that our intentions or decisions are epiphenomena, nor that our actions are predetermined. In other words, the experiments do not refute either libertarianism or compatibilism.²⁸

1.2 Technical issues

Before we delve into the many conceptual worries one (especially a philosopher) might have about Libet’s experiments, we will briefly consider some more technical criticisms of Libet’s experimental setup. These criticisms mostly focus on two aspects of his research: (1) Libet only measured activity in a specific part of the brain, and (2) Libet used a debatable method for timing the awareness of the intention to act.

²⁷This apparent advantage that compatibilists have over libertarians with regard to Libet’s results, might be the reason why some compatibilists have gleefully accepted the conclusion of Libet-style experiments. Markus Schlosser, for instance, argues that we can “altogether avoid Libet-style challenges if we adopt a traditional compatibilist account of free will” [Schlosser 2012, p. 365].

²⁸A more detailed look at libertarianism and compatibilism will follow in Chapter 2.

In Libet's experiment only a small part of the brain is monitored: the supplementary motor area (SMA). Activity in the SMA is mostly associated with the *late* stages of motor-planning, therefore Libet is overly quick in identifying readiness potential with the initiation of the spontaneous action. There might be indicators of conscious decision-making elsewhere in the brain, and it might even be the case that these indicators precede the onset of readiness potential. Indeed, in a more recent fMRI study [Lau, Rogers, Haggard *et al.* 2004] conscious intention was associated with neural activity in areas other than the motor cortex. Unfortunately these experiments did not have the time resolution to show whether or not this elsewhere located activity preceded activity in the motor cortex.²⁹ Libet might of course argue that the activity in other areas is irrelevant. As long as the readiness potential precedes the consciousness of a decision to move, the conclusion that spontaneous motor actions originate unconsciously remains upright. This reply is not fully satisfactory however. After all, the W-moment might only indicate the time at which the agent becomes aware of the moment at which she will flick her wrist, whereas the decision to flick her wrist 'fairly soon' or 'as soon as possible' is already consciously made before the readiness potential starts building up.

The aspect of Libet's experiment that has received the most scrutiny is his method for timing the conscious intention to act. Daniel Dennett and Marcel Kinsbourne [1992] note that since test subjects had to divide their attention between the CRO and their own inner mental workings, reports of the conscious decision to move might be significantly later than the decisions themselves. Their idea is that processing the fact that one has made a decision, and processing the visual information from the CRO, both take a certain amount of time. And according to them the time it takes

²⁹The diminished time resolution is one of the drawbacks of using fMRI over EEG.

to process this information might lead to a mismatch between the actual timing of the W-moment and the test-subjects reports of that timing. Furthermore, because of the relatively short time scale—everything, from the onset of readiness potential to the movement, happens in about half a second—even very small inaccuracies in the agent’s estimate of her decision time could prove fatal to Libet’s conclusions about the relative timing of conscious decision and brain activity.

Dennet and Kinsbourne’s suggestion that agents are inaccurate when it comes to estimation is backed up by a large body of neuroscientific research. Humans simply are not very good at retroactively recalling when something has happened. A study by James Moore, Daniel Wegner and Patrick Haggard [2009], for example, shows that subjects grossly overestimate the interval between their key press movement and a tone. In that particular study three of the fourteen test subjects were erratic to such a degree, that their data had to be omitted from analysis.³⁰ Even Libet himself [1983] found that agents misestimate the timing of their actual movement by about 50ms. Of course, the fact that the CRO moves very fast (a full round takes 2.56 seconds) does not help the tests subjects’ accuracy a lot either.

Apart from increasing the potential for inaccuracies, the usage of the CRO is also problematic for a different reason. Recall that the test subject has the dual task of making a decision when to move, and of monitoring his decision time on a revolving clock before him. Thoughts like ‘I will move when the dot is at 6 o’clock’ could easily slip into the subject’s mind. And these thoughts may precede the RP, so that the motoric action is not unconsciously initiated after all.³¹ Libet did, however, try to preclude earlier

³⁰For more research on this see: [Ulrich *et al.* 2006, Lau, Rogers and Passingham 2006, Sarrazin *et al.* 2008].

³¹This ties in with the worry that indicators of conscious activity in other parts of the brain might precede the readiness potential.

thoughts about the action by asking his subjects “to let the urge to act appear on its own at any time without any preplanning or concentration on when to act” [Libet *et al.* 1983, p. 625] But it is questionable how successful this strategy is. Is it really possible to monitor the time of your decision to act, without any “concentration on when to act”?

Despite these issues with Libet’s experimental methods, his results have been duplicated many times.³² And his conclusions about the timing of the movement, the readiness potential, and the awareness of decision, are widely accepted as empirical facts.³³ More importantly, even if Libet’s results concerning the timing of the respective events are sound, they still do not in the least bit threaten free will—as the next section will show.

1.3 Some conceptual issues

Even if we accept Libet’s results, it is not immediately clear what should be concluded from his data. Some have taken Libet’s result as strong indicators for the view that human free will on the whole is an illusion.

The idea that perceived freedom of action is an illusion fits in a long tradition of philosophic and psychological thinking [...]
Libet’s findings are the first direct neurophysiological evidence in support of this thesis. [Banks and Pockett 2007, p. 658]³⁴

But many philosophers have argued, on conceptual grounds, that Libet’s results do not in the least bit warrant the conclusion that free will does not exist. I have classified the conceptual worries about drawing this conclusion in three categories. First, it is argued that Libet’s results do not show that the brain *determines*

³²For example in [Keller and Heckhausen 1990] and [Haggard and Eimer 1999].

³³For a further defence of Libet’s (timing-)methods see [Klein 2002a,b,c].

³⁴See, e.g., Spence [1996] for a similar claim.

that we will act before we can decide (§1.3.1). Second, even if Libet's results show that the test-subjects were unfree in executing their simple motoric movements, these results cannot be generalised. In other words, we cannot conclude that free will on the whole is an illusion, if we find that the performance of some very specific tasks is not free (§1.3.2). Finally, Libet interprets the detected readiness potentials as the brain making a decision, but he does nothing to rule out alternative interpretations that might even be more plausible (§1.3.3).

1.3.1 Initiation and determination

Libet and his co-authors themselves were very cautious—at least in their original article—about the conclusion one should draw from their experimental results:

[A]ccepting our conclusion that spontaneous voluntary acts can be initiated unconsciously, there would remain at least two types of conditions in which conscious control could be operative. [Libet *et al.* 1983, p. 641]

One of these “types of conditions” will be discussed in this subsection, the other in the next (§1.3.2). The first situation in which conscious control could still be operative is the following:

(1) There could be a conscious ‘veto’ that aborts the performance even of the type of ‘spontaneous’ self-initiated act under study here. [ibid, p. 641]

Conscious vetoing has become widely known as ‘free won’t’. We do not need to go into the intricacies of the debate concerning this veto-power and whether it is truly free.³⁵ What is more important for the purpose of this chapter, is what the possibility of

³⁵Elisa Filevich *et al.* [2013], for instance, argue that there is no *free* won’t, because antecedent brain activity predicts our decision to inhibit simple motor action. However, this experiment is at least as conceptually dubious as the original Libet experiments.

interference after the act is initiated, tells us about the type of unconscious origination Libet is talking about. The unconscious building up of readiness potential apparently is not *sufficient* for the simple motoric action to occur. Hence it is not the case that the occurrence of readiness potential *necessitates* the occurrence of the motoric action. To the contrary, in a recent experiment where subjects had to decide either to press or not to press a key when they heard a tone, it turned out that readiness potential was building no matter whether the subjects pressed or refrained from pressing [Trevena and Miller 2010].

If the readiness potentials do not necessitate the motoric actions, then the most that Libet's experiment shows is that some events that are prior to, and might be correlated with the decision to act, do not need to be conscious in nature.³⁶ Interestingly, no plausible contemporary view of free will presupposes that these early correlates must be conscious. This is easiest to see for compatibilists. They explicitly believe that we have free will even if our actions were fully causally determined. Moreover, even the most devoted libertarians, who believe that indeterminism is required for free will and that agents should be the ultimate sources of their acts, do not require that nothing prior to an action can be correlated with it. Consider my colleague walking in and asking me to join her for lunch. This is not something I consciously control, but still it is heavily correlated with me going for lunch. Nevertheless, I could decide not to join her because I have an important deadline coming up, or because I'm on a hunger strike against the academic 'publish or perish' mentality. It is my possibility of doing so that makes me the ultimately source of my decision to go for lunch. Having conscious control over everything correlated with my de-

³⁶That is, if we are ignoring the worries expressed in the last section about the possibility of earlier indicators of conscious activity in other parts of the brain.

cision is not required.³⁷ Moreover, we do not need neuroscience to show us that many of the actions we perform have early unconscious causal correlates, as the example of me going for lunch shows. The fact that neuroscientists are attempting to show that the earliest correlates of our free actions are unconscious—as we will see John-Dylan Haynes and his colleagues try to do exactly the same thing—while no plausible philosophical theory of free will requires them to be consciously controlled, shows how hopelessly neuroscience and philosophy are talking past each other.

Philosopher Timothy O'Connor agrees that Libet's experiments do not threaten free will. Nevertheless, he argues that we should amend our philosophical conception of freedom based on neuroscientific results. In the cases where we do not interfere between the unconscious initiation of an action and its execution, our behaviour is more or less automatic. Therefore, "the freedom and responsibility of much of what we do must be thought of as "inherited" from the comparatively few directly free choices that we make" [O'Connor 2009b, p. 181].³⁸ But should we agree that these 'automatic' actions are not directly free, or less free? As long as we are aware of where an unconsciously initiated chain is leading us, we can interfere. That possibility of interference seems enough for me to be in control of my action.³⁹

This concludes the discussion of the first conceptual worry philosophers have concerning the interpretation of Libet's results as a threat to freedom. To summarise: Libet's experiments at most show that some early correlates of our actions might arise

³⁷Even so-called agent causal libertarians who think that an agent's causing of her action must itself be uncaused, do not deny that prior events can, e.g., be reasons that influence an agent's decision—albeit in a non-causal manner. See Chapter 4.

³⁸We see a similar idea in Kane [1999a] who believes that only important 'character forming actions' are directly free, all our other actions are free only derivatively.

³⁹Much more on how philosophers understand the control required for freedom in Chapter 2.

unconsciously, but this is something we already knew before the dawn of neuroscience. Furthermore, no plausible theory of free will requires that the earliest correlates of our actions do have to arise consciously.

1.3.2 Generalisation

Let me now turn to the second class of conceptual worries: worries concerning the extent to which Libet's results on simple motoric actions performed in an experimental setting are applicable to all of human agency. If Libet's experiments are supposed to prove that free will on the whole is illusory, it is important that such a generalisation is possible.⁴⁰ Otherwise it could simply be the case that we do possess free will, but that we simply do not employ this capacity of choice in the specific subset of actions Libet studied.

It should be striking to anyone who reads the original description of the experimental procedure of Libet's experiment how abnormal, how contrived, and how far removed from anything we would normally call free human action, the wrist flickings of his test subjects are.

The subject sat in a partially reclining position on a lounge chair with an observer present in the room. Each trial was started only when the subject considered himself comfortably ready. The trial began with a brief 'get-ready' tone. This signalled that during the next 1-3 s the subject should relax his muscles, especially those of the head, neck and forearm, blink his eyelids if he wished [...] Subjects were asked to maintain their gaze fixed on the centre of the CRO screen and not to follow the CRO spot around, even though they were to report information [of the] 'clock-position' [...] Subjects were trained to make their self-initiated movement sufficiently brisk so that within no more than 10-20 ms from the start of any EMG potentials they achieved the amplitude pre-set to trigger the computer. The sub-

⁴⁰Again, Libet himself did not intend his experiments to disprove free will.

ject was asked not to blink from the time the CRO spot started revolving until after the event [...] the subject was told that he may blink during the trial if the need arose; but that, if he did blink (or made some other extraneous motion), he should then simply wait for the CRO spot to make at least another full revolution before performing the quick voluntary movement [...] The subject was asked to wait for one complete revolution of the CRO spot and then, at any time thereafter when he felt like doing so, to perform the quick, abrupt flexion of the fingers and/or the wrist of his right hand [...] The subject was instructed 'to let the urge to act appear on its own at any time without any preplanning or concentration on when to act' [...] Within a few seconds after the event, the subject was asked for his report of that timing, as in recalling a spatial image of ordinary clock time in conjunction with another event. It was emphasised that only an after-the-event recall of the experience was required, and that the subject should not worry about the task in advance of each event. [Libet *et al.* 1983, pp.625–626]

Note that this entire experimental procedure only occurred after one or in some cases two half-day sessions of training. And even in this experimental situation there were sessions where some subjects did report awareness of the decision to flick the wrist *before* the readiness potential. Indeed Libet's results were that readiness potential preceded awareness of decision by 350 ms *on average*—i.e. averaged out over all subjects on all sessions. But test subject B.D. had a session consisting of 40 trials where he/she on average reported the decision about 65 ms before the readiness potential was registered. B.D. is not alone, subject eight of Haggard and Eimert's [1999] variant of the Libet experiment, for instance, reported her decision on average 454 ms before a laterlised readiness potential was measured.⁴¹ It is unclear what we should conclude from these results. Did some subjects not follow the instructions

⁴¹A lateralised readiness potential or LRP is measured in only one of the hemispheres.

correctly? Can we attribute the discrepancies to measurement errors? Perhaps subject B.D. and subject eight are the ‘chosen ones’ instilled with the powers of freedom, while the rest of us are will-less zombies? Whatever we conclude, even if Libet’s experiments only show that some of our actions—the quickly performed spontaneous simple motoric actions—are not free most of the time, this should still be concerning to those who believe in free will. Did we not use to believe that even these simple motoric actions were free all the time? And do Libet’s results, therefore, not at least show that free will is illusory for a subset of our actions? How we should respond to these questions depends on whether we really thought that the simple motoric actions Libet studied were free in the first place. Can we be sure that these actions are free given the above described abnormality of the conditions in which these actions had to be performed? Perhaps the answer should be ‘No’ if we consider the fact that participants in Libet’s experiment *were obliged* to flick their wrists in each trial. It was not allowed to just sit there in that partially reclining chair without feeling the urge to press—at least a test-run without a button press would not be considered a test-run at all.

Owen Flanagan [1992] makes a related point. He notes that there are many cases in which we make a conscious decision to do something, and then let our unconsciousness handle it from there. I can decide to ride home on my bike after completing this paragraph, and execute that action while thinking about how to formulate the next paragraph, without ever spending another thought on my actual bike ride. Flanagan claims that there are even cases where somebody decides to drive home, suffers a petit mal seizure on the way, loses consciousness, and still completes the drive. In Libet’s experiment, Flanagan argues, the only free decision is made *before* the experiment actually starts. It is made when the test subject agrees to partake in the experiment and to

follow the experimenter's instructions. Hence, it is not surprising that Libet does not find data pointing towards free conscious decisions *during* the experiment. Moreover it seems clear that the decision to partake in the experiment and follow the instructions are causally relevant to the subject's simple motoric actions. Without the decision to partake in the experiment the particular wrist flickings would have never happened. When Libet concluded that the simple motoric actions that are performed during his experiment did not originate consciously, he was simply not looking back far enough, or so Flanagan argues.⁴²

As we saw above, Libet himself is well aware of the limited scope of his results. Recall that the original article which presented the experiments claims that there might still be sets of conditions in which "conscious control could be operative." The first set of conditions obtains when we exercise our power to veto decisions our brain has made.⁴³ Here is the second set:

(2) [Conscious control can be operative in] those voluntary actions that are not 'spontaneous' and quickly performed, that is, in those in which conscious deliberation (of whether to act or of what alternative choice of action to take) precedes the act, the possibilities for conscious initiation and control would not be excluded by the present evidence. [Libet *et al.* 1983, p. 641]

But is this class of actions based on deliberation, not much more relevant to the free will debate than the more or less random actions Libet studies in his experiments? One reason to think this, is that the freedom of deliberation-based action is the kind of freedom that can ground moral responsibility—what reason a person has for doing something is important when assigning praise or blame.⁴⁴ However, as I argued in the introduction to this thesis, it is best to attempt to settle issues concerning free will inde-

⁴²See [Zhu 2003] for similar arguments.

⁴³This was discussed above (§1.3.1).

⁴⁴Adina Roskies [2011], for instance, follows this line of argument.

pendently of moral considerations. Nevertheless, there is another reason why the class of deliberative actions is more relevant to free will than the class of random selections. Champions of free will, after all, do not argue that every bodily movement is free. Knee reflexes, for example, clearly are not. They are arguing that our *willed* actions are free. But as Elizabeth Anscombe [1957, §37] convincingly argues, you cannot will or want a thing without at least finding something desirable about it. It would be nonsensical to proclaim: “I want a saucer of mud”, wait for someone to bring it to you, and then say “My want is gratified” while walking away and forgetting about it. If you want something you must have a reason for wanting it: you want the saucer of mud, because you like the feel of mud, or because you want to paint with it, or perhaps because you want to prove a philosophical point. But there is some nuance here. An action *can* be performed for a reason even if it is not *preceded by* conscious deliberations. I have good reasons for brushing my teeth every night, but I hardly ever think about these reasons just before (or while) brushing my teeth. However, if you were to ask me why I brushed my teeth I could immediately come up with a fine rationalisation.⁴⁵ Although Libet does not mention the possibility of rational actions that are not preceded by conscious deliberation, we can easily see that his experiments exclude the possibility of having any reason at all. I cannot prefer flicking my wrist right now over flicking it in three seconds because nothing is more desirable about flicking it now as opposed to flicking it a few moments later. Furthermore, because the test subjects are not allowed to deliberate or pre-plan, they cannot even try to construct faux reasons like: ‘It is more suitable to press now, because the arm of the clock before me points exactly towards my wrist.’

Hence the argument can be made that Libet’s experiments do

⁴⁵For more on the sort of consciousness required for intentional action, see Chapter 5.

not at all study the initiation of “free voluntary action”, instead they study agents that are forced to pick between two equal courses of action that they are not allowed to think about. In this light it actually seems quite a feat that humans are able to do so in the first place.⁴⁶

1.3.3 Alternative interpretations

A final conceptual issue concerns the interpretation of the results of Libet’s experiment. Is Libet correct in thinking that the readiness potentials he measures signal that the brain makes a decision to act? One problem with this interpretation is that Libet seems to think that decisions are instantaneous, but, as Dennet [2004] argues, it would be more plausible to think of decision making as a temporally extended processes. We can lean one way, then the other, decide one thing, veto that decision, and do the other thing. Libet asked his test subjects to report a single moment at which they had decided to move. But, if decisions aren’t instantaneous it could be that the subjects only reported the conclusion of their decision. If this is correct, then it seems possible that the ramping up of readiness potential merely corresponds to an early phase of conscious decision making, instead of to an unconscious decision made by the brain. As we have seen, Libet tried to preclude this by asking his subjects to refrain from preplanning and to try to decide spontaneously, but to what extent could you call something that happens only 0.35 seconds before the action “pre-planning”? At least it seems that this alternative interpretation of Libet’s cannot be ruled out directly.

Note that I have so far mixed up the usage of the words ‘intention’, ‘wanting’, ‘urge’, and even ‘decision’. In doing so I have followed Libet and the majority of neuroscientific literature on this

⁴⁶In section 1.4, I will argue that disproving free will by means of Libet-style experiments is impossible because they necessarily have to study random selection or tie-breaking, instead of intentional action.

topic. However, Alfred Mele [2007, 2009] has correctly emphasised the need for conceptual rigour in this context. He argues that there are important differences between wanting to do, or feeling the urge to do something on the one hand, and deciding to do something on the other. “For example, you might have an urge to scream at an irritating colleague but decide not to” [Mele 2007, p. 190]. Now, Mele suggests that the readiness potential does not correspond to a decision to act made by the brain before the subject becomes aware of it. Instead he believes that it corresponds to a proximal urge of the agent.⁴⁷ The fact that urges or wantings might arise unconsciously is not at all detrimental to free will—we can after all refrain from acting on the urge. The idea that readiness potential signals an urge rather than a decision in fact meshes well with Libet’s finding of the veto power. We feel the urge to scream at the colleague (the readiness potential builds up) and we decide not to (we veto). Furthermore, I believe that Mele’s view finds further support in actual neuroscientific research. Section 1.2 mentions the experiment by Judy Trevena and Jeff Miller [2010] where subjects had to decide to either press or not press a key when they heard a tone. The result was that readiness potential was building up no matter what the subjects ultimately decided. It could now be hypothesised that in this experiment the tone caused the subjects to feel an urge to press and that that is why the potential was building up. Even more striking is the following result by Itzhak Fried:

As part of a clinical procedure to identify which brain areas might be particularly involved in [...] epileptic activity, [we] stimulated a range of brain areas directly via electrodes and asked the patients to report what

⁴⁷Of course, it might be dubious in the first place to relate *particular* happenings in the brain (like the building up of readiness potential) to *particular* mental events (like decisions, wantings or urges). But I will ignore such worries here.

they felt. When they stimulated the SMA [Supplementary Motor Area], some patients felt an urge to perform a movement or anticipated that a movement was about to start. When the stimulating current was gradually increased, the patients made movements of the same body part that they previously wanted to move. [Fried *et al.* 1991]

The SMA is exactly the area where Libet measured readiness potentials and Fried's research confirms that electrical activity in this area is related to urges, lending further support to Mele's suggestion. We can conclude, in any case, that Libet was too quick in supposing that the measured readiness potentials correspond to the brain's making a decision. Plausible alternative interpretations are available, and Libet's experiments do not tell us much about free will before these interpretations are sorted out.

This concludes my discussion of the conceptual worries concerning Libet-type EEG experiments.⁴⁸ Although I have not discussed every conceptual worry that has been raised in the vast amount of literature that reacts to Libet's experiments, I believe to have addressed the most important issues. In the next section we will look at a more recent Libet-style experiment using fMRI techniques. But before we do so, let me list the main takeaways from this section:

1. Libet's experiments do not show that our brain decides to act before we do, it merely shows that the earliest causal correlates of our actions might arise unconsciously.

⁴⁸I will not extensively discuss the experiments reported in [Keller and Heckhausen 1990] and [Haggard and Eimer 1999]. Although these experiments improve upon Libet's original design and also distinguish between generalised and lateralised readiness potential, the main conceptual worries expressed in this section directly carry over to these experiments. For an intricate critical assessment of the particularities of these experiments, see Bayne [2011].

2. It is unlikely that Libet's results concerning the simple motor actions he studied in his experimental setting, can be generalised to all human actions. Even stronger, it is unclear whether we should classify the movements he studies as free actions at all.
3. Because Libet is conceptually inaccurate, alternative interpretations of his data are possible.

It seems that each of these issues separately already presents a conclusive case against the thesis that Libet's experiments disprove the existence of free will. But if you take them together it is absolutely clear that Libet's experiments are a threat neither to compatibilism nor to libertarianism. The experiments do not show that conscious decisions are epiphenomena that play no causal role in the action production, neither do they establish that unconscious brain events determine our actions.

1.4 Haynes's experiment

Despite the overwhelming amount of criticism on the suggestion that Libet's experiments prove that free will is an illusion, free will denying neuroscientists are currently as vocal as ever in the public debate. When confronted with the conceptual issues surrounding Libet's findings, they are eager to point out the outdatedness of Libet's methods and chastise philosophers for being so far out of touch with contemporary science that they have still come no further than discussing 30 year old research.⁴⁹ By now, they claim, Libet's results have been strengthened and radicalised, especially by John Dylan Haynes's research group who found predictive neural activity seven(!) seconds before a voluntary choice occurred [Soon *et al.* 2008].

⁴⁹For an example see Lamme [Lamme 2011b], Alfred Mele reports similar experiences with neuroscientists [Mele 2014].

The critique on the slow pace of philosophy is not fully unwarranted. Indeed, not much work has been published yet on Haynes's results. In this section I hope to fill this lacuna. I will argue that although Haynes's experimental design improves quite a bit upon Libet's, most of the *conceptual* issues carry over. In other words, (1) Haynes's experiments merely show that there are early correlates of our actions we are unconscious of, (2) It is unlikely that Haynes's results can be generalised to the type of action that we normally have in mind when we think about free actions, and (3) alternative interpretations of Haynes's data are possible.

The authors of [Soon *et al.* 2008] point out three shortcomings of Libet's original experiment and explain how their methods are superior. Their improvements all address possible flaws in Libet's experimental design (we discussed these in section 1.2). Interestingly, any serious consideration of the pressing conceptual worries expressed in the last section is strikingly absent from both this article as well as their supplementary discussion. Like Libet, then, Haynes and his colleagues are mainly concerned with identifying neural correlates of decisions prior to the time at which test subjects report awareness of the wanting, intention or urge to press a button.

The first improvement on Libet's original experiment is that subjects could choose between two different buttons. In the Libet experiment subjects only chose *when* to act, but in Haynes's experiment subjects also had to decide on *what* to do.⁵⁰ Providing the subjects with multiple behavioural options is primarily done to rule out the idea that early brain activity merely reflects unspecific preparatory activation.⁵¹ But Haynes [2011] seems to think that providing multiple buttons also takes the actions performed in his

⁵⁰This improvement is also implemented in the experiment by Haggard and Eimer [1999].

⁵¹As is suggested with regards to readiness potentials by the earlier mentioned experiments of Trevena and Miller [2010].

experiments closer to real-world actions, and hence contributes to the likelihood that his experimental results can be generalised to actions performed outside of his experimental setting. We will discuss whether this indeed is the case below (§1.5.3). First let us look at the rest of the experimental design.

The second difference between Libet's experiment and that of Haynes's group concerns the timing-method. Subjects did not have to remember the clock position of a moving dot on a CRO. Instead they were instructed to look at a screen where consonants were presented, one at a time for 500ms each. Subjects had to passively observe the letter-stream, and right after they pressed a button, they had to report which consonant was present when they made the conscious decision to press. This method of timing has two advantages over Libet's method. First, because the letters were displayed in a random order the sequence was unpredictable which avoids possible pre-planning on part of the agent.⁵² Furthermore, the letter method avoids the inaccuracies that can occur with rotating stimuli like Libet's CRO.⁵³ A drawback of Haynes's timing method is that its time resolution is a lot coarser than that of Libet's method. Every consonant is on the screen for half a second, therefore, if an agent reports that she decided to press when e.g., a *P* was on-screen, this only reveals that she decided somewhere between, say, 500ms to 1000ms before the button press. However, this coarse-grainedness of the time resolution is not a large issue given the fact that predictive brain activity was found 7 seconds before the subject's report of her decision.

The final and most important difference between the two ex-

⁵²Whether this aim is truly achieved is questionable. Since only eight different consonants were shown, subjects could of course still have thoughts like: "I will press the left button the next time a *P* comes around". However, further measures to ensure the spontaneity of the decision were taken by means of explicitly instructing the subjects not to pre-plan.

⁵³Discussed in §1.2.

periments is that the latter uses fMRI instead of EEG. 'fMRI' stands for functional magnetic resonance imaging. It is a method to measure the degree of activation of specific brain areas by means of detecting blood flow. The higher the level of activation of a brain region, the more oxygen it needs. Oxygen is carried to brain cells by the haemoglobin in blood. Since active neurons need more oxygen than inactive neurons, blood releases oxygen to them at a greater rate. Therefore, the ratio between oxyhaemoglobin (haemoglobin that has bound an oxygen molecule) and deoxyhaemoglobin (haemoglobin without the bound oxygen), differs in different brain areas depending on their activation. The fMRI scanner can pick up on this because oxyhaemoglobin and deoxyhaemoglobin differ in magnetic susceptibility. Statistical methods are used to translate the raw data into the familiar multi-coloured brain scan pictures. A drawback of fMRI, is that it has a very coarse grained time resolution.⁵⁴ In Haynes's experiment only one image is generated per every 2 seconds of activity. But again given the nature of his results, the large temporal resolution of his fMRI measurements is, just like the large resolution of his timing method, no hindrance. The biggest advantage of the fMRI method is that Haynes and his collaborators were not limited to measuring activity only in the SMA. Instead they measured activity in a wide selection of different brain areas, thus overcoming the criticism levelled against Libet that he could not exclude there being preparatory activity in the brain occurring before his measured readiness potentials.⁵⁵ In fact, earlier activity is precisely what they found. Furthermore, where the EEG only measures an *increase* of activity, Haynes was able to detect shifts in the *patterns* of activity in certain areas even when there was no change in overall activity.

Haynes and his colleagues analysed the fMRI images they ob-

⁵⁴For more on the technical reasons for this limitation see Kim *et al.* [1997].

⁵⁵See §1.2.

tained to discover when and where they could find brain activity that was predictive of the subject's decision to press either left or right. For their analysis they used pattern recognition algorithms. First, such an algorithm has to be trained: roughly, this works by feeding it fMRI images, telling it the time at which these images were taken, and telling it whether the images were taken from a trial in which the subject pressed the left button or from a trial where the right button was pressed. After training, the recognition algorithm is fed a new image and is told the time at which it was taken, but not whether it is a picture from a left or a right trial. The algorithm is then asked to 'predict' whether the subject pressed left or right. This means that it compares the new image to the training images, and determines whether it looks more like an image of a left, or of a right trial. That procedure was repeated multiple times for different brain areas, different time points and different test subjects. In this way Haynes and his colleagues were able to determine the predictive accuracy of activity in different brain areas at different times.⁵⁶

As I already mentioned above, the results were striking. Haynes found predictive information up to seven seconds before the agent reported a decision. This information was found in the fMRI signals from one particular area of the frontal polar cortex: the Brodmann area 10 (BA10).⁵⁷ The predictive accuracy of the information was 60%, which means that the algorithm was 10% better than chance at predicting whether the left or the right button would be pressed. The 'predictive' accuracy of information from the motor cortex after the reported choice reached 75%. A further result was that the time at which the button press occurs

⁵⁶For a more technical explanation of the pattern classification techniques used in [Soon *et al.* 2008] see their supplementary methods section and [Haynes 2011].

⁵⁷Because it takes some time for the oxygen rich blood to reach the active brain area, this can mean that there is predictive brain activity up to 10 seconds before the subject consciously decides which button to press.

could also be “significantly predicted” by activity in the SMA and pre-SMA about 5 seconds before the time of decision.⁵⁸

1.5 Conceptual worries about Haynes

Do the results described in the previous section tell us anything about the existence of free will? Haynes thinks they do. He writes:

[O]ur findings do address one specific intuition regarding free will, that is, the naive folk-psychological intuition that at the time when we make a decision, the outcome of this decision is free in the sense of not being predetermined by prior brain activity. [Haynes 2011, p. 16]

Compatibilists deny that free will requires such a lack of pre-determination. Hence Haynes’s results form no threat to their conception of free will and some compatibilists even opportunistically cite Haynes’s experiment in favour of their view.⁵⁹ But should libertarians really be worried? If Haynes is right that the outcome of our decisions is predetermined then this would certainly cast doubt on their indeterministic conception of free will. However, I will argue in this section that Haynes’s experiment does not warrant the conclusion that our decisions are predetermined. Many of the same conceptual worries raised against Libet’s experiments (or better: raised against concluding that free will is an illusion on the basis of Libet’s experiments), can also be applied to concluding that free will is an illusion from Haynes’s results. In the next section I will argue that, in fact, no experiment in the style of Haynes or Libet can ever produce evidence that free will does not exist.

⁵⁸See page 21 of the supplementary methods of [Soon *et al.* 2008].

⁵⁹See, for instance, [Schlosser 2012].

1.5.1 Haynes and alternative interpretations

In one of the few philosophical responses to Haynes's experiment, Alfred Mele maintains that his alternative interpretation criticism of Libet carries over to Haynes's experiment [Mele 2014]. Recall that Mele stressed the importance of distinguishing between an urge to move and a decision to move. One can have the urge to flick one's wrists while ultimately deciding not to.⁶⁰ Haynes and his colleagues are sloppy, Mele argues, in claiming that the early activity in BA10 "encoded the subject's motor decision ahead of time" [Soon *et al.* 2008, p. 544]. For, according to Mele, a decision cannot be encoded until it is made. And hence, Mele concludes, the predictive activity does not necessarily predetermine that the subject will make a certain decision, it might merely indicate that the subject has a certain *inclination* to either press the left or the right button.⁶¹ Moreover, "a person who moves on from having a mere inclination to do something to actually deciding to do it has made significant progress toward doing it" [Mele 2014, p. 302].

Unfortunately, Mele's psychological interpretation of the early activity in BA10 as corresponding to an inclination is a lot less plausible than his hypothesis that readiness potentials correspond to urges.⁶² For subjects in Haynes's experiment were asked to act immediately after they became aware of their urge or decision to do so. Therefore it is unlikely that brain activity that happens *seven seconds* before the agents reports awareness of a decision, corresponds to an inclination felt by the subject. Nevertheless,

⁶⁰In §1.3.3 mentioned Mele's suggestion that a measured readiness potential corresponds to an urge felt by the agent rather than to her 'brain making a decision'.

⁶¹I think Mele is here reading too much into the word 'encoded'; as far as I can see Haynes's and his colleagues use 'encoded' simply in the technical sense of 'is predictive of'. Nevertheless, Mele is right that we cannot conclude that specific neuronal activity predetermines decision, from the bare fact that it is predictive of decision. More on this below (§1.5.2).

⁶²And unlike that hypothesis, the idea that the BA10 encodes inclination is not backed up by experimental results.

even if the early activity in BA10 does not signal a conscious inclination, as Mele thinks, this does not rule out that there might be other possible interpretations of that activity.⁶³ Therefore, more research is needed to sort out precisely how we should interpret the patterns Haynes’s experiment picks up on. For now, let us proceed to the other conceptual issues.⁶⁴

1.5.2 Haynes and determination

In section 1.3.1 we discussed that although Libet’s readiness potentials might be *necessarily* occurring before simple motoric action, their occurrence might not be *sufficient* for the motoric act to also come about. And therefore, readiness potential does not *determine* decision. In the case of the predictive activity in BA10, it is even clearer that this activity does not necessitate a particular choice. This simply follows from the fact that the predictive accuracy of that activity was merely 60%.⁶⁵ Even if activity occurred that seemed predictive of a left button press, the right button was eventually pressed in 4 out of 10 times. In the face of this it is puzzling that Haynes seems to think that his experiment shows that our decisions are not free because they are “predetermined by prior brain activity” [Haynes 2011, p. 16]. What Haynes’s remark shows, I think, is that the concept of determination he employs is

⁶³One possibility is that such activity is related to mental option-generation. For an enlightening discussion of this phenomenon, see Annemarie Kalis *et al.* [2013].

⁶⁴Mele also argues that Haynes’s results actually nicely confer to his preferred version of libertarianism, i.e. event-causal libertarianism. As we will see in Chapter 2, event-causal libertarians explicitly endorse that different choices have different objective probabilities of being caused by the agent’s reasons. Mele suggests that Haynes’s experiment is picking up on these objective probabilities. But this is dubious, because, as will be discussed below (§1.5.3 and §1.6), the choices performed in Haynes’s experiment aren’t based on reasons. I will argue against event-causal libertarianism in Chapter 2 and Chapter 3.

⁶⁵Slightly higher percentages were reached in a later experiment [Bode *et al.* 2011].

radically different from the concept of determination under discussion in the philosophical free will debate. When libertarian philosophers claim that free will is incompatible with our action's being predetermined, the determinism they have in mind is *metaphysical determinism*. Metaphysical determinism is the theory that there is only one possible future. Such determinism is a threat to free will, libertarians argue, because it implies that long before I was born it was already fixed (or determined) that I would write this sentence. Hence, I had no control over writing the sentence, and thus could not have been free in writing it.⁶⁶ Contrast this form of determination with the way in which brain activity in BA10 determines a right or left button press. Haynes's experiment is focused on unearthing the unconscious *determinants* of free actions. The term 'determinant' is most often used in the medical sciences. *A* is a determinant of *B*, if *A* is a factor that enhances the probability of *B*. Such determinants are also commonly referred to as 'risk factors'. Smoking, for instance, is a determinant for lung-cancer. But clearly, smoking does not predetermine cancer in the strong sense of metaphysical determination, i.e. in leaving open only one, cancer-ridden, possible future. Smoking at most makes developing cancer more likely. Similarly, the right-predictive activity pattern in the BA10 at most 'predetermines' a right button press in the sense that it makes it more likely that the subject ends up pressing right. As Adina Roskies remarks: "[P]redictability is at best a poor cousin to determinism, and one that can betray its familial roots" [Roskies 2010, p. 112].

Although we do not know why the early brain activity is predictive of the test-subject's decision, let's suppose it is because such activity exerts a small causal influence on the decision to press a specific button. How much should this worry us about the existence of free will? Not at all. There are numerous influ-

⁶⁶This is a very rough version of the consequence argument, which will be more extensively discussed in §2.1.

ences on our actions that we are unaware of and perhaps Haynes's experiment has discovered one more influencing factor. But the only thing needed for free will is that there are situations where these influences are not so strong that they only leave open one possible future.⁶⁷ If the low predictive accuracy of the activity in BA10 discovered by Haynes's experiment indicates one thing, it is that the occurrence of such activity indeed leaves open multiple different behaviours.

Moreover, we should ask how striking the fact that Haynes can predict motoric actions seven seconds before they happen are. Certainly the advanced fMRI and pattern recognition techniques employed by Haynes's research group are awe inspiring, but aren't we often able to make more accurate predictions about what a person is going to do way more than seven seconds in advance? Because it is a cold day, I can reliably predict that my colleague will put on his coat before he comes to work in two hours, even though he is currently sound asleep and unaware of the cold.⁶⁸ That I can quite accurately make such predictions, however, does not tell us anything about my colleague's free will. He might, just as well, decide to not come to the office today.

But what if neuroscientists could eventually reach absolute predictive accuracy? Would it not be a threat to free will if they could correctly predict behaviour all the time? Perhaps, but can such a high accuracy ever be reached? In his 'Decoding and predicting intentions' [2011], Haynes discusses why a predictive accuracy of 100% has not been reached. One reason for the inaccuracy,

⁶⁷As I remarked in the introduction, a prisoner is still free to choose from the options she has, even though these options may be more limited in that they do not include leaving the correctional facility.

⁶⁸This idea is not new, Dennett [1981, pp. 65–66] once described a behaviour prediction competition between an Earthling who uses folk-psychology and a Martian (the neuroscientist in our story) who employs the physical stance. In order to be just as successful as the Earthling, the Martian has "to avail himself of much more information about an extraordinary number of interactions of which, so far as he can tell, the Earthling is entirely ignorant."

he argues, is that there are technical limitations on our ability to measure neural signals.⁶⁹ But he admits that these limitations “cannot be used to argue that one day with better methods the prediction will be perfect; this would constitute a mere “promissory” prediction” [Haynes 2011, p. 17]. Instead, it might be the case that the early neuronal patterns in the BA10 are only partially predictive. In that case, he writes, “the signals have the form of a biasing signal that influences the decision to a degree, but additional influences at later time points might still play a role in shaping the decision” [ibid, p. 17].⁷⁰ In the light of these observations, Haynes argues that both technical limitations and incomplete determination are possible causes of the lack of full predictive accuracy. To me it seems unlikely that full predictive accuracy could ever be reached by improving our methods. After all, we often make choices quickly in reaction to external stimuli, hence it is impossible that all our choices are perfectly predictable 7 seconds before we make them.⁷¹ Furthermore, it is in principle possible to allow feed-back of real-time predictions of future decisions to the test subject.⁷² But would it not be strange, when the agent finds himself unable to press the right button, if he sees that a left press is being predicted? Therefore it is likely that, although the predictive accuracy might be increased by improving

⁶⁹He mentions that the spatial and temporal resolution of fMRI “only reflect a strongly impoverished version of the information” [Haynes 2011, p. 16], and that his decoding method would ideally need a very large or infinite number of training samples; however, “the slow sampling rate of fMRI imposes limitations on the training information available” [ibid, p. 17].

⁷⁰It is unclear whether this bias plays a *causal* role, (i.e., whether the early signals’ influence on decision is a causal influence), since “even a complete 100% prediction may not directly imply a causal link between the early predictive signals and the choice” [ibid, p. 17].

⁷¹The possibility of quick decision making also forms a problem for Libet’s experiments. As Tim Bayne [2011] points out, we sometimes make choices in way less than 500 ms, which is the time between RP and movement, making it impossible that every movement is preceded by such an early RP.

⁷²Haynes calls developing methods that allow real-time feedback “an important future line of research” [Haynes 2011, p. 9].

the methods of measurement, the early signals in BA10 simply are not fully predictive and do not fix what we are going to do. Again, this is strangely at odds with Haynes's claim that his findings reject the "naive folk psychological intuition" that our decisions are not predetermined. Clearly that naive intuition, if it really exists, is about predetermination in the strong metaphysical sense. Haynes's results at most imply predetermination in the same sense in which smoking 'predetermines' cancer.

1.5.3 Haynes and generalisation

The final conceptual worry expressed in section 1.3 was the issue of generalisation. Can we use the results on the specific kind of actions studied in Libet's and Haynes's laboratories to draw conclusions about human agency in general? That Haynes's results cannot hold for all human actions is already clear from the fact, mentioned above, that some actions are decided on and performed in way less than seven seconds. It is also clear that Haynes's experimental situation, like Libet's, is pretty remote from everyday action contexts. Again, subjects were asked not to pre-plan their actions while paying attention to the time of their decision, and again it was stressed that the time of movement was completely up to them, although they had to press either left or right in each trial—refraining was not an option. However, the worry about generalisability is the only conceptual worry of which Haynes and his colleagues take note, albeit implicitly. They see it as a limitation of Libet's experiment that it only offers the agent a choice of *when* to act. Therefore they increase the number of degrees of freedom by offering an additional 'what choice', that is by offering the choice between pressing a left or a right button. In [K. Smith 2011] there is talk of another, yet to be published Libet-style experiment. In this experiment the test subjects were asked to add or subtract two numbers from a series being presented on a

screen, and Haynes is quoted as claiming that “deciding whether to add or subtract reflects a more complex intention than that of whether to push a button, and [...] that it is a more realistic model for everyday decisions” [K. Smith 2011, p. 24]. But how much does adding complexity to the decisions in the experiment help Haynes? Not much, because no matter how complex the tasks get, and no matter how many of these tasks can be chosen, they still do not constitute more than a case of mere tie-breaking between alternatives that do not matter to the agent. The choices the test subject has to make are still arbitrary choices. And, as Roskies remarks: “Arbitrary action is, at best, a degenerate case of freedom of the will, one in which what matters about freedom fails to hold” [Roskies 2011, p. 18].⁷³ There simply is no reason to prefer the one task over the other. And, just as in Libet’s experiment, the fact that pre-planning (deliberation) is strictly prohibited, means that agents cannot even try to construct a reason to prefer one of the options. Therefore the actions Haynes’s studies, just like the actions Libet studied⁷⁴, might not be free in the first place. At least they are very unlike the actions we normally think of when we talk about free will, since those actions are performed for a reason.

We can conclude that, on the whole, Haynes’s experiment suffers from the same conceptual issues as Libet’s experiment: it is unclear how to interpret the detected patterns of brain activity, such brain activity does not determine the decision, and finally, there are reasons to suspect that the results obtained do not hold for less contrived actions. Because of these conceptual issues Haynes’s experiment does not present convincing evidence

⁷³Haynes seems to admit this when he says that his and Libet’s experiment “do not address real world decisions that have high motivational importance” [Haynes 2011, p. 16].

⁷⁴See §1.3.2.

against the existence of free will. In the next section I will go one step further and argue that experiments in the style of Libet and Haynes can never be successfully employed to disprove free will.

1.6 The futility of Libet-style experiments

In this section I will expand on the worry about the possibility of generalising Haynes's and Libet's results (§1.5.3 and §1.3.2) to argue that Libet-style experiments can never disprove free will.⁷⁵ With 'Libet-style experiment' I mean experiments that correlate a decision to engage in spontaneous motor activity with brain activity that occurs before the subject reports conscious awareness of the decision. As we have seen, both Libet and Haynes study subjects who have to select between equal options without pre-planning. Apparently human beings are able to perform these random selection tasks—we are not like Buridan's ass who dies of starvation because he cannot decide between two equally appealing stacks of hay. But how do we do it? Maybe we sometimes use environmental stimuli as a tie-breaker, but in Libet's and Haynes's experiments the environment is always the same. Another option would be that we use random fluctuations in cortical activity. And it could be hypothesised that these fluctuations are the ones that Haynes's pattern-recognition algorithm is picking up on. Similarly, John Eccles [1985] argues that Libet's readiness potentials reflect such spontaneous fluctuations in cortical activity. No doubt, more research on random selection is needed to find out how we really do it. However that may be, it is clear that such random selection is different from the action selection on the basis of reasons that we normally associate with free will.⁷⁶

⁷⁵To be clear, I am not claiming that free will cannot be the object of experimental investigation at all, I am merely suggesting that Libet-style experiments cannot teach us much about freedom.

⁷⁶I am not suggesting that Haynes and Libet have singled out a type of unfree action, as Adina Roskies [2011] seems to think, and that their results therefore

It might be tempting to conclude from this, that Haynes and his fellow neuroscientists could start to study reason-based actions by devising Libet-style experiments in which they offer non-arbitrary alternatives. However, there is a fatal problem with this: the arbitrariness of the decisions studied is crucial to the experimental setup of Libet-style experiments. First, in order to assess how much better than chance the prediction algorithm is, it is important that the decision made by the test subject really is a fifty-fifty decision. The neuroscientist can, of course, not ask his test subjects to balance their decisions, because that would entail that they have to keep track of earlier decisions and pre-plan future ones. Haynes and his colleagues instead attempted to ensure that the decisions made during the experiment were really arbitrary by testing several subjects the day before the real experiment and selecting only those subjects that spontaneously balanced their decisions.⁷⁷ Second, let us suppose that test subjects were presented with a slightly non-arbitrary choice between say, a chocolate chip cookie and an Oreo. Now, how much of a result would it be if neuroscientists would detect activation patterns based on which they are able to predict, with a certain accuracy, that subject *A* will pick a chocolate chip cookie, before *A* herself reports that she has made that decision? They might have just picked up on subject *A*'s love for chocolate chip cookies, well known among her friends and family. Our preferences are bound to show up in our brain somehow. But it would be odd to suggest that subject *A*'s preferences decided that she would have a chocolate chip cookie

contradict our intuition that we are free to select from random alternatives. Even in tie-breaking situations we can still control which option we select: it is not the case that we decide on one option and then surprise ourselves by picking the other option. And, as discussed above (§1.5.2), the fact that our selections can be predicted, does not mean that they are pre-determined (in the metaphysical sense).

⁷⁷It could be smugly remarked that test subjects who used their free will to, say, consistently press the right button, were not even allowed to partake in the experiment.

before she herself could make this decision. The more plausible story would of course be, that she herself made her decision *based* on her preferences.⁷⁸

Hence there seems an inescapable dilemma for the neuroscientist who attempts to disprove free will by means of a Libet-style experiment. The first option for the neuroscientist is that she studies arbitrary non-deliberative decisions. This will make sure that the unconscious determinants she finds have nothing to do with the agent's reasons for acting. But it has the downside that she would be studying tie-breaking instead of volitional action. The second option for the neuroscientist is to study reason-based action. This makes it more likely that she is actually studying what philosophers think of as free action. But then she could not claim that these action are unfree by detecting early predictive activation patterns, because these patterns could merely indicate the prior reasons on which the agent acted. Therefore we should not only conclude that the Libet-style experiments conducted thus far have failed to disprove free will, we should conclude that it is impossible in principle to reject free will on the basis of Libet-style experiments.

1.7 Priming and illusionary willing

In the previous section I have argued that it is impossible to disprove the existence of free will by means of Libet-style experiments. Although these experiments play an important role in the arguments of the free will sceptic, other scientific results are also cited as 'evidence' against free will. This section will look at the most prominent of these results.

One experimental psychologist who has been beating the drum

⁷⁸Things are further complicated because these preferences are not fixed—after having had ten chocolate chip cookies in a row, even subject *A* might start to be more inclined towards selecting an Oreo next.

of free will skepticism for some time, is John Bargh. He writes:

There is no need to posit the existence of free will in order to explain the generation of behavioral impulses, and there is no need to posit free will in order to explain how those (unconscious) impulses are sorted out and integrated to produce human behavior and the other higher mental processes. The phenomenological feeling of free will is very real, just as real for those scientists who argue against its actual existence as for everyone else, but this strong feeling is an illusion, just as much as we experience the sun moving through the sky, when in fact it is we who are doing the moving. [Bargh 2008, pp.148–149]

In support of this conclusion, Bargh mentions a host of behaviours that are causally produced in ways we are not aware of, like our unconscious mimicry of other's behaviours and the behaviours in his own experiments on so-called *psychological priming*. In one of his priming experiments Bargh presented subjects with a scrambled-sentence task. As a matter of fact, some of the subjects were presented with a task that contained a disproportionate amounts of words like “worried, Florida, old, lonely, grey, selfishly, careful, sentimental, wise, stubborn, courteous, bingo, withdraw, forgetful, retired, wrinkle, rigid, traditional, bitter, obedient, conservative, knits, dependent, ancient, helpless, gullible, cautious, and alone” [Bargh *et al.* 1996, p. 236]. What Bargh found out, was that subjects who were ‘primed’ in this way with the ‘elderly-stereotype’ tended to walk slower than a control group who was not primed. But why would Bargh think that this threatens free will? Di Nucci [2012] suggests that Bargh might think priming limits free will because, had a subject not been primed, she would have walked faster. And that therefore the subject ‘has been made’ to walk slower in such a way that her control over her own behaviour has been diminished or taken away. But as Di Nucci convincingly argues, this is clear nonsense. After all, the subject could have walked faster had she wanted to. If she had to catch a train, or if

the fire alarm went off, or even if she suddenly felt like it, she would have had no problem with sprinting away.⁷⁹ The real reason, then, why Bargh thinks that his experiment does have any implication for the existence of free will, could be that he assumes that all of our behaviours can be fully explained by things like priming effects. The running in reaction to the sound of the fire alarm, he might think, is also just caused by further psychological, sociological or evolutionary priming. But why would we suppose such a thing? One of the characteristics of most priming studies is that the behaviour that is caused by the priming is behaviour the test subject is unaware of. Why would we even expect that the agent would control this behaviour? Of course agents cannot consciously bring about their unconscious behaviours. But these results on unconscious priming effects do not tell us anything about consciously performed intentional actions. It seems that Bargh can only view his results as undermining free will, because he already assumes free will scepticism from the start.

Another result cited by free will sceptics in support of their position is that agents sometimes come up with erroneous backward looking rationalisations.⁸⁰ This phenomenon is studied in one of the most cited experiments in the history of psychology.⁸¹ Participants were asked to choose one pair from among four identical pairs of stockings. It was found out that most participants had a clear preference for the right most pair. When asked why they chose that pair, however, none of the participants cited its position relative to the other pairs. Instead they came up with a fabricated rationalisation and cited small relative differences in colour

⁷⁹In a similar experiment, subjects were primed with the ‘helpful’ stereotype and hence showed a tendency to pick up the experimenter’s pen when it fell down. Interestingly, this effect was immediately counteracted if the pen was visibly leaking. See [Macrae and L. Johnston 1998].

⁸⁰For Victor Lamme [2011a] the existence of a rationalising mechanism, the “babble box” as he calls it, seems his main reason for his denial of free will

⁸¹See [Nisbett and T. D. Wilson 1977].

or quality of the fabric. These rationalisations, of course, were obviously false, since the experimenters made sure that the stockings were exactly identical. But why would this experiment be an argument against the existence of free will? Perhaps our conception of free will would be in danger, if we were to find out that *all* of our expressed reasons for acting are pure nonsense. But can we conclude that from the fact that we sometimes come up with invented rationalisations in cases where we are asked for a reason and do not want to disappoint the experimenter? Drawing that conclusion is akin to drawing the conclusion that we can never come to know anything, because our senses are sometimes misled by optical illusions. Hence, the free will sceptic who claimed to provide evidence for his doubt of the existence of free will, turns out to be a radical sceptic from the very start.

A clear example of someone who presupposes, rather than argues for free will scepticism is Daniel Wegner [2002, 2008]. He discusses a large amount of examples of what he calls ‘the illusion of conscious will’. His cases vary from superstitious sports fans who believe that their behaviour can influence the outcome of a game, to an experiment where a test subject is made to believe that someone else’s arms are his own. What all the cases have in common is that a subject thinks that she is in control of something, when in fact she is not. Hence, Wegner concludes that the experience of controlling an action does not necessarily indicate that the action is consciously controlled. So far this is correct, but then he suddenly leaps to the conclusion that our experience of conscious control is *always* illusory. That, of course, is just plain scepticism. To be fair, Wegner does cite one further reason for his sceptical conclusion, albeit not a very strong one: the cases of illusions of conscious willing show that “the brain structure that provides the experience of will is separate from the brain source of action” [Wegner 2002, p. 47]. But as Eddy Nahmias [2002]

correctly points out, the fact that these brain systems are not identical does not mean that they are unconnected.

Finally, cases of neurological disorders or brain damage are often presented as evidence against the existence of free will. One example is the alien hand syndrome, a disorder in which a person experiences one of her hands as operating with a mind of its own. Another is the case of Kenneth Parks who unknowingly killed his mother in law, while sleepwalking. In both of these cases the patients's brains do cause their actions, but they do so in the absence of an experience of conscious willing. Hence, these cases are the opposites of Wegner's illusions—in those cases there was willing, but no action. But to conclude from these cases of neurological disorder, that all of our actions must be unfree, again amounts to mere scepticism. In fact, it would be even stranger to draw such a conclusion from these cases, since there evidently is something wrong with the subjects in them: they have a disorder. We shouldn't conclude that no one can see from the fact that some people are blind, and we shouldn't conclude that no one has a capacity of choice, because some people have disorders that diminish it.

1.8 Concluding remarks

In this chapter we have seen that (neuro-)scientists do not make a convincing case against free will. I have argued that Libet-style experiments cannot be used to prove the non-existence of freedom, and showed how other results can only be seen as a threat to free will if one already presupposes free will scepticism. In the light of this we might wonder where such scepticism comes from. My hypothesis is that one of the causes of scepticism is a rather widespread misunderstanding about the nature of free will. Consider what Professor Read Montague, director of the Human Neuroimaging Lab at the Virginia Tech Carilion Research

Institute, writes:

Free will is the idea that we make choices and have thoughts independent of anything remotely resembling a physical process. Free will is the close cousin to the idea of the soul, the concept that “you”, your thoughts and feelings, derive from an entity that is separate and distinct from the physical mechanisms that make up your body. From this perspective, your choices are not caused by physical events, but instead emerge wholly formed from somewhere indescribable and outside the purview of physical descriptions. This implies that free will cannot have evolved by natural selection, as that would place it directly in a stream of causally connected events. [Montague 2008, p. 584]

I would be sceptical about free will as well if I really believed it would imply such ethereal entities. We all should be. Unfortunately, Montague is not the only one with a distorted picture of free will. As Mele [2014] shows, there are many neuroscientists who confuse the existence of free will with the doctrine of substance dualism. But even many neuroscientists who do not believe that free will requires the existence of extra-natural entities, still believe that free will would require that our actions originate in uncaused or spontaneous neural events we are conscious of. It is because of this presupposition that Libet’s and Haynes’s experiments are supposed to disprove the existence of free will. For they show that our actions can be causally correlated to earlier unconscious events. But do we even know what a brain event that would arise instantaneously, is conscious, and is fully separate from all prior neural events, would look like?

In any case, there is no reason to deny the existence of free will because these kinds of events do not exist. Hence Dennett calls Wegner “a killjoy scientist who shows that Cupid doesn’t shoot arrows and then insists on entitling his book *the Illusion of Romantic Love*” [Dennett 2004, p. 224].

But contemporary philosophers are not fully free from blame when it comes to the conceptual confusion either. Most of them also believe that all of our actions are preceded by mental events that have neural correlates.⁸² Indeed, the most commonly held theory of action states that the difference between mere behaviour and intentional action consists in the latter being caused by preceding mental events.⁸³ But, as mentioned above, we often find that we are doing something without prior deliberation, desiring or intending. This morning I got out of bed, dressed, had breakfast, brushed my teeth and rode my bike to the office, all the while thinking about this chapter. Does this mean that I was not conscious of doing these things? Well, I was not consciously thinking about them while doing them, but if someone had stopped me and asked me why I was riding my bike, I would not for a moment have been surprised to find myself riding that bicycle, and would have immediately been able to reply something like “because I’m going to the office”. As I will argue in Chapter 5, the kind of consciousness signalled by my absence of surprise, is sufficient for *control*: although I was thinking about my thesis I was aware that I was riding my bike, and if my thoughts had made me realise that I had accidentally left my notes at home, I would have been perfectly able to turn my bike around to go and get them.

How we should understand the control we have over our actions is the topic of the following chapters. For now we should conclude that neuroscience has failed to disprove free will, because it is confused about the nature of its target. I could therefore not agree more with Daniel Dennett when he writes:

If you are one of those who think that free will is only *really* free will if it springs from an immaterial soul that hovers happily in

⁸²Although most philosophers would deny that free will implies that these events should arise spontaneously.

⁸³The first version of this causal theory of action (CTA) was developed by Donald Davidson [1963]. CTA will be extensively criticised in Chapter 3.

your brain, shooting arrows of decision into your motor cortex, then, given what *you* mean by free will, my view is that there is no free will at all. [Dennett 2004, p. 222]

But unfortunately Dennett continues:

If, on the other hand, you think free will might be morally important without being supernatural, then my view is that free will is indeed real, but not quite what you probably thought it was.

With this latter statement, I could not disagree more. Free will is exactly what we thought it was: the capacity to choose between *real* alternatives that are open to us.

Chapter 2

Free will and Control

If it can be shown that actions are not completely constituted by behavioral episodes with a certain kind of special mental causal history, compatibilism will be in trouble.

—John Bishop⁸⁴

Philosophers have been occupied with the study of free will for over two millennia. But they no longer monopolise the topic. As we saw in the previous chapter, neuroscientists have recently tried to end the philosophical discussion once and for all by claiming that free will is nothing but an illusion. But, as we saw, neuroscience does not even come close to vindicating that grotesque statement. The main reason why neuroscientists think that free will does not exist is that they have a hopelessly erroneous view of what free will must be. They believe that free will consists in the conscious origination of brain events that are entirely uncorrelated to anything that is prior to their origination. As I argued, this extreme form of conscious origination is not required for free will. What is required is that an agent has (conscious) *control* over

⁸⁴Bishop [1989, p. 66].

her action. Not coincidentally perhaps, much of the contemporary free will debate is centred around the concept of control. Philosophers ask what kind of control free will requires, and how much of such control is needed for free action. In this vast literature, many different specifications of control have been offered, and in this chapter we will take a look at some of the proposals.

Even so, I will not provide a comparative analysis of the merits and flaws of the different accounts of control. Instead I will try to show that many of the accounts, despite their differences, also share a central idea: they construe free action as intentional action+ X . In other words, they think of free actions as intentional actions that fulfil certain further conditions. Interestingly, the contemporary discussion of free will focuses almost entirely on these purported extra conditions that turn a merely intentional action into a free action. Hence, the main way compatibilists and libertarians contest each other is by proposing different sorts of extra features that either are or are not compatible with determinism. In this chapter I will argue that the libertarian advocate of the simple picture of free will should reject this way of conducting the debate. Instead she should not argue that determinism is incompatible with any extra feature that distinguishes free action from merely intentional action, but that it is incompatible with intentional action itself.

This chapter starts with a sketch of the contemporary philosophical debate about free will in section 2.1. We will see that both compatibilists and libertarians are faced with the problem of explaining how control is possible on their accounts, i.e., they have to explain how we can control our actions in deterministic or indeterministic scenarios respectively. As I will point out, there is a surprising amount of agreement between compatibilists and most libertarians about what such agential control consists in: control consists in the causation of action by the reasons of the agent. In

section 2.2 I will argue that, although this causal account helps both the libertarian and the compatibilist to explain how agents can control their actions, it also removes the fundamental grounds they have for disagreeing with each other. In other words, as soon as one accepts the causal account of control one can no longer hold on to the idea that determinism or indeterminism is required for free will. In section 2.3 we will see that both libertarians and compatibilists attempt to specify extra conditions on agential control to justify their respective requirements of indeterminism or determinism. But I will argue that as soon as the libertarian starts playing this game, she has already lost. Instead the advocate of the simple picture of free will should deny that agential control can be accounted for in terms of causation by reasons. She should deny that there is any form of agential control compatible with determinism. As section 2.4 points out, this implies that she should not just deny that free action is incompatible with determinism, she should deny that intentional action as such is possible in a deterministic world.

2.1 The free will problem: a problem about control

Determinism is the thesis that what will happen in the future is fixed. Once, scholars might have thought that the future is determined because God already knows what will happen. But for the more materialistically inclined, the thesis of determinism might simply be the idea that everything that will happen is settled by what has already happened in conjunction with the laws of nature. Although it might seem to us that the future is one of myriad possibilities, for someone who “at a certain moment would know all forces that set nature in motion, and all positions of all items of which nature is composed, [...] nothing would be uncertain

and the future just like the past would be present before its eyes” [Laplace 1814/2007][p. 4]. Hence, determinism implies that our idea of the open future is only an illusion that results from our hampered epistemic positions—we are no Laplacian demons.

But does determinism also imply that free will is an illusion? *Prima facie* there certainly is friction between the idea of choice as selecting one among many alternatives, and the deterministic contention that there is but one alternative. Indeed, as Peter van Inwagen points out:

Once, one might have said that the problem of free will and determinism—in those days one would have said ‘liberty and necessity’—was the problem of discovering whether the human will is free or whether its productions are governed by strict causal necessity. [van Inwagen 1986, p. 1]

Liberty and necessity, freedom and determinism, were seen as strict opposites. And in fact many non-philosophers still think that way. John Dylan Haynes, one of the main protagonists of the previous chapter, for instance, says that the intuition that we have free will “is scientifically implausible anyway, simply because it stands in contradiction to our belief in deterministic laws of physics” [Haynes 2011, p. 16].⁸⁵ However, most contemporary philosophers are compatibilists. They believe that the perceived tension between free will and determinism is just that: only an apparent tension that has no foothold in reality. Some of them, let us call them *weak compatibilists*, believe that the question of (in)determinism does not pertain at all to the question of free will. Others, *strong compatibilists*, go even further and believe that free will *entails* determinism.

In the face of these widespread compatibilisms it becomes question begging to claim that free will is not reconcilable with determinism simply because liberty and necessity seem to be opposites.

⁸⁵A strange remark anyway since contemporary physics recognises indeterministic laws.

Therefore, the contemporary philosopher of free will has to do a little more work to bring out what problem determinism presents to free will. Perhaps the best attempt at doing so is van Inwagen's consequence argument:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. [van Inwagen 1986, p. 16]⁸⁶

What this argument brings out, is that a deterministic world view makes it hard to comprehend how we can still be the authors of our own actions, how our actions can be “up to us”. However, the sense in which free will requires that our actions are up to us is, not surprisingly, itself a point of philosophical contention. Hence, compatibilism can overcome the consequence argument, if it shows that there is a relevant sense in which an agent's acts can be up to her even when the occurrence of these acts is already determined. In other words, any good compatibilist theory of free will at least has to (1) offer an account of how agents *control* their own actions, and (2) show that this kind of control is not threatened by determinism.

Explaining how agents control their actions is not only of importance for compatibilists, it is equally important for those who believe that we have a kind of free will that is incompatible with determinism. If free will is only problematic in the face of determinism, one could wonder why the free will problem persists given the widespread scientific belief in indeterminism that has arisen since the dawn of quantum mechanics. Finding out that the world is not deterministic by itself does not help much in understanding

⁸⁶I will not here go into van Inwagen's more formal formulation and the possible technical problems with that version of the consequence argument. For a good discussion of these issues see Chapter 1 of O'Connor [2000].

how our actions can be up to us. In fact, indeterminism might make it even harder to understand how we can exert control over our own actions. Indeterminism, some would argue, merely injects randomness into the world. Hence, if human action is indeterministic, it just seems to be a mere matter of chance, similar to the outcome of a dice roll. And if there is anything we cannot control, it is the outcome of a chancy process. To illustrate this “problem about luck for libertarians”, Al Mele [2006] asks us to imagine two worlds, W_1 and W_2 , that have the exact same laws of nature and up till now have had the exact same past. Now consider Joe, a free agent that decides to A in world W_1 . According to the libertarian, Joe could decide not to A in world W_2 . The libertarian, after all, does not believe that the past and the laws of nature fix current decisions. But, Mele complains:

If there is nothing about Joe’s powers, capacities, states of mind, moral character, and the like in either world that accounts for this difference, then the difference seems to be just a matter of luck. [ibid, p. 9]

If our choices are just matters of luck, then what we decide is up to chance, and hence cannot be up to us. Thus an indeterministic world view makes it hard to understand how we can be the authors of our own actions, how we can control what we do. Hence the libertarian is faced with a task similar to that of the compatibilist. Any good libertarian theory of free will also has to (1) offer an account of how agents control their own actions, and (2) show that that kind of control is not threatened by *indeterminism*.

Together, the consequence argument and the luck problem present what we could call ‘the problem of control’ or simply ‘the free will problem’. Conjunctively the consequence argument and luck problem seem to show that, no matter whether the universe is deterministic or indeterministic, we cannot control our actions and hence do not possess free will.

The problem of control can be used to frame the most prominent positions in the free will debate. Libertarians are convinced by the consequence argument, but believe that we have free will because we can escape the luck problem. Strong compatibilists are convinced by the luck problem, but believe that we have free will because the consequence argument can be answered. Weak compatibilists believe that neither the consequence argument, nor the luck problem are destructive of free will. Finally, there are hard incompatibilists who think that both the consequence argument and the luck problem are insurmountable and therefore believe that we have no free will.⁸⁷ For someone new to the free will debate, this might seem to sum up all the available positions quite nicely. But unfortunately this is not the case. Not only are there many variants of both libertarianism and compatibilism,⁸⁸ there also are hybrid positions. Some philosophers, for instance, claim that they are agnostic about the compatibility of freedom and determinism, yet do believe that moral responsibility is compatible with determinism, giving rise to such a monstrosity as ‘semi-compatibilism’.⁸⁹ And it seems to be only a matter of time before someone will claim the opposite and defend that we cannot have responsibility in a deterministic world, but that freedom is perfectly attainable in it. Perhaps testament to how cluttered up the free will debate has gotten, is Seth Shabo’s recent attempt to turn Peter van Inwagen’s [2000] statement that he cannot see how to escape the luck objection and that free will therefore remains mysterious to him, into a full fledged philosophical position called ‘mysterianism’.⁹⁰

⁸⁷Derk Pereboom is the main advocate of this position. See, e.g., Pereboom [2002, 2007].

⁸⁸The Stanford encyclopedia, for example, lists more than ten different variants of compatibilism, all with their own further divisions and subdivisions. See, McKenna [2009].

⁸⁹John Martin Fisher and Mark Ravizza [2000] are the main proponents of such a position.

⁹⁰See, Shabo [2011, 2013]. Colin McGinn [1989] has famously defended a

This is not the place to go into the details of all these positions and to compare their respective merits and flaws. To do so would require another dissertation, or two. But it is interesting to note the enormous explosion of positions we have had in the last few decades. Perhaps part of this is just a consequence of the increased publication pressure which awakens the philosopher's insatiable appetite to invent, occupy and vigorously defend her own position. But there might be another reason for philosophy's failure to settle on just a few main contenders in the free will debate. The explosion of positions might also be due to there being something fundamentally wrong with the way the debate is conducted and with the questions that are discussed. For when the topic of discussion is unclear, positions tend to be all over the place. Now I want to avoid making too grand a historical claim about these recent developments in the history of the philosophy of free will, but I do think that there is something very odd about the current debate.

Above, we saw that both compatibilist and libertarian accounts of free will have to fulfil requirement (1): they have to offer an account of how an agent controls her actions. Given the fact that they draw opposite conclusions about the compatibility of control and determinism from such an account, it is striking that almost all compatibilists and a vast majority of libertarians accept the very same account of agential control, i.e., Donald Davidson's [1963] Causal Theory of Action (CTA).⁹¹ Perhaps the mutual acceptance of CTA would not be so striking if questions about the compatibility of CTA with determinism and indeterminism were huge topics of contentions in the current debate. But these questions are almost never asked. In fact it seems quite clear that CTA

similar mysterianism about consciousness.

⁹¹The libertarian minority that does not *fully* accept CTA is made up of so-called non-causal libertarians and agent-causal libertarians. Their positions will be discussed in Chapter 4.

is compatible with both determinism and indeterminism.

CTA says that an agent controls her actions iff they are caused by the agent's reasons, but it does not say anything about whether this causal relation needs to be deterministic or indeterministic. Hence, if both parties in the current free will debate are accepting CTA we can start to wonder what it is these philosophers are arguing about. Well, the response might be, although both libertarians and compatibilists accept CTA, they do not believe that CTA provides the full story that has to be told about free action. What CTA explains, is only how we should distinguish between intentional action and mere bodily movement. Free will theorists should also distinguish between free and un-free intentional action. They want to know what transforms "a mere action into a free action" [Franklin 2011a, p. 203]. Hence, although CTA provides the agent with a certain degree of control over her own actions, such control is insufficient for freedom. The control required for freedom needs an extra ingredient, it consists of intentionality+ X as I put it above. Hence what philosophers in the free will debate are arguing about is this all important X -factor. And it might be that the extra X is exactly what makes freedom incompatible with determinism or indeterminism.

Indeed, the extra X is the main topic of discussion in the free will debate, but this, I believe, renders much of the debate superficial. Contemporary libertarians, for instance, stress that we need indeterminism, or alternative real possibilities in addition to the causal story about agential control. But as I will show next, to escape the luck problem libertarians don't need that extra factor, they only need CTA. However, since CTA also allows the compatibilist to rebut the consequence argument, it makes little sense for this libertarian to keep insisting on indeterminism. As soon as the libertarian accepts the causal theory of action, I argue, she buys into an inherently compatibilist doctrine and thus has

already lost the debate.⁹²

2.2 CTA and the problems of control

Event-causal libertarianism is the account of free will that combines libertarianism with the causal theory of action. It is by far the most common brand of libertarianism found in the current debate. Main proponents include Robert Kane [1985, 1996, 1999a,b] and Mark Balaguer [2004, 2010]. But we will start this section by looking at what I consider to be the clearest formulation of this position provided by Christopher Evan Franklin [2011a, 2014]. The main merit of Franklin’s account is that he is especially clear about how the libertarian employs CTA to escape the luck problem.

At first sight, it might seem that this problem is not too hard to escape. Consider Mele’s free agent Joe.⁹³ In one world Joe decides to *A*, and in the other he decides not to *A*. Now even though there might be nothing “about Joe’s powers, capacities, states of mind, moral character, and the like” that explains this difference, we do have one explanation for why Joe *A*’d in the one world and did not do so in the other. That is, he *decided* to *A* in the one world and decided not to *A* in the other. Mele is right that nothing can explain the occurrence or non-occurrence of *A* other than Joe’s decision. But this is no argument against the libertarian, since it is exactly her position that we have this undetermined capacity to shape the future according to our own decisions.⁹⁴ Although this may be an adequate reply to Mele, it does not fully repel the luck problem. To bring this out we should look at perhaps the most poignant explication of the worry that

⁹²Strictly speaking, CTA is inherently *weak* compatibilist. Since CTA provides an answer to the luck problem, as soon as the strong compatibilist accepts CTA, she is robbed of the major reason to insist on determinism.

⁹³See p. 68.

⁹⁴Randolph Clarke offers a reply similar to this. See Clarke [2003, pp. 77–82].

indeterminism precludes freedom: van Inwagen's [2000] rollback argument.

Van Inwagen considers Alice, a free agent who is torn between telling the truth and lying, but in the end tells the truth. If libertarians are right, Alice could also have lied. Therefore, if God were to roll back the universe to the point right before Alice told the truth, and let things proceed again, she might end up lying. Imagine that God does this a large number of times, say a thousand, then we would get a certain ratio of, say, three hundred lies to seven hundred truth tellings. But then it seems that there just is a 70% probability that Alice tells the truth and a 30% chance that she lies. However, if Alice's actions are just a matter of chance, then it does not seem to be something over which she has any control, i.e., her original telling of the truth was no free action after all.

According to van Inwagen, if what Alice does is up to chance, it would even be wrong to say of Alice that she is *able* to tell the truth, or *able* to lie. For if Alice knew that the ground floor probability of her telling the truth was only 70%, then she could no longer sincerely promise that she will tell the truth. Of course her making of the promise to tell the truth could increase the probability that she will, but can you really promise something if you know that there is even a small chance that you will not hold it? Of course you can, if you believe that something external has a chance to interfere with your ability to carry out the promise, but that is not what we are considering here. We are considering the chance that you will, out of your own accord, decide not to keep the promise, even if all other things remain equal. Now if Alice is not even able to promise someone that she will tell the truth, she would feel that what she will end up deciding, is beyond what she can *control*. And it is because luck threatens control that it threatens freedom.

Mele argued that Joe's actions were a matter of luck, because nothing could explain why he decided to *A* in one world, and not to *A* in the other. We replied that Joe's decision might explain the difference. But with Alice the case is not so simple. The problem van Inwagen sketches is not that we lack an explanation of why Alice does one thing or the other: it directly threatens Alice's ability to control what she does by arguing that what she does is a mere matter of objective probability.

Van Inwagen himself is disheartened by this result. He expresses the hope and belief that something is wrong with it, but cannot pinpoint the error. However, Franklin [2011a] claims that he can. He analyses the rollback argument (and any other variant of the luck problem) as consisting of two basic steps:

- (1) If human actions are undetermined, then they are a matter of chance.
- (2) If actions are matter of chance then they are not free.

According to Franklin (1) is obviously correct. "All undetermined actions have an objective probability of less than 1 of occurring and so all their occurrences are a matter of chance" [ibid, p. 216].⁹⁵ But he disagrees with (2). The fact that human actions are a matter of chance does not imply that we have no control over them, and hence that we are unfree.

Franklin is especially unconvinced by van Inwagen's argument that luck strips us of the ability to lie or tell the truth. Alice supposedly is not able to tell the truth because she knows that even if she decides to tell the truth she still might end up not doing so. Thus Alice finds herself in a situation where there are certain objective probabilities concerning what she will do at a later time and there is nothing she can do to affect these probabilities. Franklin deems such a scenario very atypical. He thinks that a libertarian agent has ways to ensure that she acts in a certain

⁹⁵Lara Buchak [2013] has recently argued against (1).

manner at a later time. An agent might have chosen at $t - 2$ to tell the truth at t and this earlier choice might have raised her probability at $t - 1$ of keeping silent at t to 1.⁹⁶ According to Franklin, van Inwagen cannot see a way out of the luck problem because he “mistakenly locates indeterminism after choice and for this reason sees indeterminism as a threat to freedom and responsibility” [Franklin 2011a, p. 217].⁹⁷ But if there is no indeterminism located after choice, Franklin argues, the agent does not have to be lucky to end up doing what she decided to do. It is not indeterminism that opens up libertarianism to the luck objection, it is the *location* of indeterminism in particular libertarian accounts that does. Therefore Franklin thinks that it is no coincidence that most formulations of the luck objection are targeted at Robert Kane’s event-causal libertarianism. Robert Kane, Franklin notes, explicitly locates indeterminism after choice. On Kane’s account it is indeterminate whether a particular effort of the will leads to action. An agent who makes an effort of the will, thus has to wait and see whether it results in an action. She can do nothing to ensure it will, and hence how she will act is up to luck.⁹⁸ But, Franklin argues, event-causal libertarians do not have to accept Kane’s account of the location of indeterminism, and therefore are able to escape the luck problem.

But why does Franklin think that indeterminism *before* choice

⁹⁶I don’t think that it is correct to think that agents are ever able to raise the probability of their doing something to 1 at any point before they act, because a free agent might always reconsider, or there might be interfering factors. But for the sake of argument, I will ignore such considerations.

⁹⁷I am not sure that this is entirely fair to van Inwagen, since, in his formulation of the rollback argument, he does not mention the location of indeterminism or Alice’s choice, instead “God caused the universe to revert to precisely its state one minute before Alice *told* the truth” [van Inwagen 2000, p. 14, my italics].

⁹⁸Sometimes the luck problem is formulated in terms of ensurance. E.g., Ishtiyaque Haji [1999, 2000] pushes the worry that if an agent’s acts are undetermined she can do nothing to ensure or guarantee that she will perform a particular action.

(or effort of the will) does not threaten control? It might no longer be a matter of chance what the agent will *do* after she has chosen, but if it remains a mere matter of chance what the agent will *choose*, how can that choice be up to her?

This is where CTA comes into play. According to CTA, intentional actions are distinguished from mere bodily movements (like the knee reflex) by their causal history. Intentional actions are events that are caused, in the right way, by reasons.⁹⁹ Reasons are thought of as agent involving mental events. Proponents of CTA differ in what kind of mental events are supposed to do the causing (beliefs, desires, commitments, intentions) but the details do not matter for our, or for Franklin's purposes. What is important, is that on this account an agent's exercise of *control* over her actions consists in her actions being appropriately caused by her mental states. An agent does not control the movement of her lower leg if it is caused by a doctor who strikes her patellar ligament with a reflex hammer, but she does control the movement if it is caused by her desire to kick the doctor. Event-causal libertarians add to this story by claiming that the causation involved in the production of action is non-deterministic causation, thus allegedly providing agents with the coveted opportunity to do otherwise. If this view is right, then deciding just is a matter of non-deterministic causation. The decision to tell the truth *consists* in Alice's reasons for truth telling causing her action. But Alice could also have decided to lie, since there was an objective probability that Alice's reasons for lying had become causally active instead of her reasons for telling the truth.

We are now in a position to understand why Franklin thinks that the location of indeterminism in the causal chain leading up to the action matters. If it is the case that even *after* we make a decision to tell the truth there still is a big chance of us lying,

⁹⁹More on 'the right way' in section 3.3.

then it seems that we have no control over our actions. However, if the indeterminism is located *before* choice then this merely means that chance determines which choice is made, i.e., which set of reasons causes the action. But whatever Alice ends up doing, it will always be *her* reasons—either those for, or those against telling the truth—that do the causing. Since CTA tells us that the exercise of agential control just consist in an agent’s reasons causing her actions, Alice is in full control whatever she ends up doing. Hence, what she does is up to her.

Markus Schlosser [2014] deems this reply to the rollback argument insufficient. According to him, Franklin has shown that luck does not threaten Alice’s ability to tell the truth or to tell a lie, but that it remains a matter of luck which ability she will exercise. Schlosser argues that Alice “lacks the power or control to exercise either one of the two abilities such that she can select which alternative to pursue” [ibid, p. 380]. But this complaint betrays a failure to comprehend the role CTA plays in Franklin’s event-causal libertarianism. CTA is a reductive account of agency: intentional action *just is* bodily movement caused by relevant mental states.¹⁰⁰ And, similarly, deciding to tell the truth rather than to lie just comes down to one set of reasons causing the action rather than another set of reasons. If CTA is the correct story of agency, there is nothing more to controlling which ability one will exercise than simply exercising one ability instead of another, on the basis of one’s reasons for exercising it. We must conclude that Franklin has successfully shown that if CTA is the correct story about agential control, then there is no ground for the claim that an agent’s action is outside of her control only because it is a matter of chance which set of reasons will cause her action.

But not all is well for the libertarian. Accepting the causal theory of action comes at a large cost: it offers compatibilists a

¹⁰⁰More on the reductive nature of CTA in Chapter 3.

way to reply to *their* problem of control, i.e., the consequence argument. Recall that argument:

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. [van Inwagen 1986, p. 16]

With the help of CTA, the compatibilist can object that we *can* have control over our current actions even if they are already settled by the past and the laws. As John Bishop argues:

[T]he Consequence Argument is false because the core claim of the Causal Theory of Action is true. The remote past and the laws of nature will indeed be un-avoidable—beyond the agent’s control. But if the agent’s behavior, which is the *deductive* consequence of these unavoidable states of affairs, is also the *causal* consequence of the right kind of states of the agent, the agent’s behavior will fulfil conditions sufficient for it to count as action. If facts about the remote past cause present behavior via the right sort of causal chain, the causal consequence of what is un-avoidable will actually constitute a case of something that comes about through agent-control. [Bishop 1989, p. 57]¹⁰¹

If CTA correctly construes agential control in terms of the agent’s reasons causing her action, then this does not depend on whether the causation is deterministic or not. Consider Bo, who inhabits a deterministic world. It has always been determined, even when Bo was not yet born, that today at six, she will go for a drink. Now the consequence argument questions whether Bo’s going for a

¹⁰¹Bishop claims no originality for this line of arguing against the consequence argument but attributes it to Michael Slote [1982]. However, it is Bishop himself who points out the centrality of CTA for compatibilism. In sections 2.4–5 of his book he even defends his opinion that “a positive case for the falsity of [the consequence argument] can be achieved only by affirming a suitable version of CTA” [Bishop 1989, p. 59].

drink is really up to her. The causal action theorist will argue that this depends on the causal route that leads up to Bo's action. If Bo is remotely controlled by an evil neuroscientist, who steers her to the bar, then her going for a drink was not up to her. But if, on the other hand, her going to the bar was caused by her desire for refreshment then Bo did control her action. After all Bo's exercise of control *consists* in her reasons causing her action.

The lesson to draw from this is that as soon as the libertarian accepts CTA, she is left wondering why she thought that freedom required indeterminism in the first place. Similarly, the strong compatibilist is left wondering why she insisted on determinism once she understands that CTA can also be employed by libertarians—as Franklin shows. CTA deals with both the luck problem and the consequence argument in one fell swoop. It explains how agents can control their actions in both deterministic and indeterministic universes. Once we accept CTA, it seems that we no longer should care about determinism or indeterminism. CTA merely reduces control to causation but says nothing about whether this causation has to be deterministic or not. Hence anyone who endorses CTA should settle on the position I have labelled weak compatibilism above: she should believe that the question of (in)determinism does not pertain to the question of free will.

2.3 Free will as intentionality+X

In the last section we saw how CTA is an inherently (weak) compatibilist doctrine that undermines the libertarian's reasons for insisting on indeterminism. Franklin himself seems to be aware of this worry when he writes: “it seems that the *only* difference between [event-causal libertarianism] and compatibilism is that the former require[s] indeterminism” [Franklin 2011b, p. 688]. However, Franklin thinks that there is good reason for this requirement. To understand what justification he thinks remains for in-

sisting on indeterminism, it helps to contrast Franklin's position with another variant of event-causal libertarianism: the position Randolph Clarke [2003, p. 57] has labelled 'deliberative libertarianism'.¹⁰²

Deliberative libertarians also endorse CTA and they also locate indeterminism before choice in order to avoid the luck objection. They differ from Franklin's position in that they do not believe that the causal links between the agent's reasons or considerations and her actions are indeterminate. What is indeterminate, on their account, is which considerations come to the mind of a deliberating agent. An agent who decided to go straight home after work, could also have decided to take a detour past the store, but only if she had considered that, say, she was almost out of cat food. Deliberative libertarians believe that their account is resistant to the luck objection because we do not normally assume that an agent has control over what considerations pop into her mind. Therefore by locating indeterminism at the time of deliberation they do not diminish the control the agent would have had otherwise—i.e., if her considerations were determined. Nevertheless, because it is indeterminate which considerations enter the agent's mind, what the agent does is not fixed by the past and the laws of nature. Hence, the promise of deliberative libertarianism thus is that it secures the ability to do otherwise without introducing luck.

Franklin is not satisfied with this position.

Deliberative libertarianism [...] is a near-sighted solution to the luck argument. By locating indeterminism during the deliberative process, they may avoid the possibility of indeterminism diminishing control, but only at the expense of indeterminism's exclusion from being relevant to enhancing control. [Franklin 2011a, p. 207]

¹⁰²The position is also known as 'modest soft libertarianism' [Mele 2006].

According to Franklin, libertarians must not only show how they can evade the luck objection, they must also provide an adequate answer to “the problem of enhanced control” [Franklin 2011b]. In other words, they must not only show that indeterminism is *harmless*, they must show how it is *beneficial* to freedom.

On behalf of the deliberative libertarian, Al Mele tries to answer this challenge. He concedes that deliberative libertarianism might not secure more control than compatibilist accounts of free will. However, he believes that indeterministic freedom might still be worth wanting:

Wilma [Mele’s spokesperson for deliberative libertarianism] observes that independence is among the things that some people intrinsically value. Some people value independence, in some measure, from other people and from institutions. Wilma values, as well, a measure of independence from the past. She values, she says, a kind of independent agency that includes the power to make a special kind of contribution to some of her actions and to her world-contributions that are not themselves ultimately deterministically caused products of the state of the universe in the distant past [...] The kind of agency she hopes for, Wilma says, would render her decisions and actions personally more meaningful from the perspective of her own system of values than they would otherwise be. Although Wilma emphasises that this kind of agency is essential to the kind of meaningful life she prizes, she says that she is not claiming that it is required for freedom or moral responsibility. [Mele 2006, p. 100]

In other words, the deliberative libertarian admits that there can be freedom in a deterministic world, but she believes that a more *valuable* kind of freedom can only be had if the universe is indeterministic. To me Mele’s defence seems hopelessly vacuous. He suggests that the value of indeterministic freedom lies in that it affords agents with independence from the past. But indeterminism just is the thesis that the future is not fixed by, and thus has some degree of independence from the past. Hence, the only reason Mele

gives to think that indeterministic freedom is more desirable than its deterministic counterpart is: “because it is indeterministic”.

Franklin is more worried about Wilma’s admission that indeterminism is, strictly speaking, not required for freedom:

Deliberative “libertarianism” is no version of libertarianism at all. Libertarianism entails incompatibilism about determinism and free will and moral responsibility, yet Mele makes it clear that deliberative libertarianism has no such entailment [...] Deliberative libertarianism, consequently, is not an avenue of response for libertarians: it is a concession of defeat. [Franklin 2011a, p. 208]

This assessment undoubtedly is correct. It is also in conformity with the last section where I hypothesised that any libertarian theory that endorses CTA would collapse into a weak compatibilism. As Franklin brings out, deliberative libertarianism explicitly does so.

After all these criticism of deliberative libertarianism we could expect that Franklin himself proposes a more satisfactory answer to the problem of enhanced control. But surprisingly, his solution is not much better than that of the deliberative libertarians. Franklin argues that event-causal libertarians can secure more control than compatibilists because indeterminism allows for more *opportunities*. Compatibilist agents only have the opportunity to do what they end up doing, whereas incompatibilist agents have the opportunity to do something else. But so far this just seems a trivial restatement of the difference between determinism and indeterminism: the one allows for one possible future, the other for multiple. Nevertheless Franklin stresses that the opportunity to do otherwise is a “significant addition” [Franklin 2011b, p. 704]:

It affords agents with the opportunity to direct their lives in more than one way, to author how their lives unfold, and to choose from among several causally open options, thereby taking a stand on the kind of person they will become. [ibid, p. 704]

This is hardly convincing. The compatibilist can easily respond that the control she offers is already more than sufficient for agents to be the author of their own lives. On her account agents act for their own reasons too, and even though compatibilist agents cannot choose from several causally open options, they will pick from among options that they believe are open to them and hence they do take a stance on the kind of person they will become. Furthermore, Franklin's reply to the problem of enhanced control is just as vacuous as Mele's suggestion that indeterministic freedom is more valuable. Indeterminism, after all, just is the thesis that more than one possible future is open to the agent, i.e. that she has more than one opportunity. Hence, Franklin's answer to the question why indeterminism enhances control simply comes down to: "because of indeterminism." We can conclude that Franklin fails to show *how* indeterminism enhances control.¹⁰³

But even if he was able to answer that question, he would still need an argument to show that *only* agents who possess such enhanced control are free. After all, libertarians believe that freedom is incompatible with determinism. Hence the question becomes why the extra control provided by indeterminism, if it indeed provides enhanced control, is the final ingredient that frees the will-less. Why is the control provided by CTA not already sufficient for freedom? In answer to the latter question, Franklin cannot point to the consequence argument—as the last section showed—because if CTA provides enough control for the libertarian to escape luck, it most certainly provides sufficient control for the compatibilist to escape that argument. Without an answer to the problem of enhanced control and without a good argument against compatibilism, event-causal libertarians cannot hold on to

¹⁰³Cf. Jacobs and O'Connor [2013, p. 179]: "While agents, on [the event-causal libertarian] account, do not have any less control over what they do than agents in a corresponding deterministic scenario, they also do not have more".

the indeterminism requirement and their position collapses into weak compatibilism.

At this point the prospects for the libertarian may seem bleak. But, I want to argue, there is a way for the libertarian to escape the grasp of (weak) compatibilism. The problem with Franklin's and other event-causal libertarian accounts, is that they view indeterminism as an *extra ingredient* that turns intentional action into free action. But what they fail to notice is that the quest for such an extra ingredient is a typically compatibilist enterprise. Frankfurtians, for instance, argue that freedom does not only require that an agent can act on her own desires, i.e., that she has agential control, but also requires that the agent has the correct second order desires, i.e., that she actually wants what she wants.¹⁰⁴ Another example are those compatibilists who base themselves on Fisher and Ravizza's [2000] idea of reasons responsiveness. Their theory is that a free agent should not only be caused to act by her own reasons, it also places a further requirement on these reasons. They need to be open to revision if the agent had been in a different situation. I will not go into the specifics of this, or any other compatibilist account right here, but merely note that they all seem to agree that freedom requires more than just agential control. It requires an extra factor in addition to intentional control. In other words, the compatibilist construes free will as intentionality+X, where the X stands for reasons responsiveness, relevant second order desires, or any other factor thought to turn merely intentional action into free action.

Now, for the compatibilist it is quite a natural move to search for an extra factor that characterises free action. After all, she believes that actions have a place similar to that of any other event in the long deterministic causal chain that makes up the history of the universe. But why would the libertarian play along? For

¹⁰⁴See, Frankfurt [1969].

it is her intuition that something special happens whenever an agent acts: the agent selects one of multiple possibilities that are left open by the universe. And the libertarian can argue that all the control required for freedom just consists in that ability to shape the future. The compatibilist, on the other hand, can have no such control because she does not allow for multiple real possibilities, let alone for an agent who has the power to select one of them. Therefore, the libertarian should not believe that determinism *diminishes* control (or that indeterminism enhances it), she should believe determinism *destroys* control.

We can conclude that libertarians who want to defend the simple picture of free will should not accept CTA and then tack on indeterminism as an afterthought. CTA is inherently compatibilist and therefore it should be attacked by the libertarian. As Bishop recognises: “there may be potential joy for incompatibilists in any argument that *undermines the whole basis of a Causal Theory of Action altogether*” [Bishop 1989, p. 66]. But according to Bishop that joy would be short-lived. For libertarians are not just concerned with attacking compatibilism, they also have to account for the possibility of action in an indeterministic universe. And Bishop sees CTA as the only way of accounting for this possibility:

[I]ncompatibilist libertarians will be burning their bridges if they seek to undermine the CTA itself in order to establish that action is impossible in a deterministic world. They will be undermining the very theory that offers the best prospects for defending the natural possibility of action at all! [ibid, p. 66]

These then are the main tasks for anyone who wants to understand the simple picture of free will: (1) undermine CTA and (2) offer an alternative account of agency that explains the possibility of action in our indeterministic universe. The first task will be the focus of the next chapter. Chapter 5 and 6 will be concerned with the latter

task. Right now we should note that if my mission statement for libertarians is correct, the focus of the free will debate needs to shift quite radically.

2.4 Free action as intentional action

Currently the philosophical debate on free will is largely independent from debates in action theory. Perhaps most thinkers who concern themselves with free will are like Bishop and simply cannot see how any account of agency other than CTA could possibly work. But even those who might have doubts about CTA generally proceed on the assumption that we must have a rough idea or perhaps an intuitive grasp of what is meant with ‘action’ or ‘intentional action’. They believe that their task as philosophers of free will lies not in clarifying the concept of action—their task is to identify the extra X that transforms mere action into free action. I have argued that libertarians who want to defend the simple picture of free will should not tag along in the search for the extra X , because indeterminism is not something that enhances agential control, it is something that is essentially required for it. Therefore free will should not be discussed independently of action theory. To the contrary, the understanding of intentional action should be the main focus of libertarians.

But if we believe that determinism destroys, instead of diminishes, control, then do we not lose the important distinction between free actions and merely intentional actions? After all, libertarians believe that indeterministic action is free, but if indeterminism is already required for intentional action, then it seems that unfree yet intentional actions are impossible. Am I really suggesting that all intentional actions are free? Clearly there seem to be situations where an agent acts intentionally but not freely. Consider someone who, at gunpoint, reads a favourable statement about a terrorist organisation that is holding her hostage. The

hostage clearly acts intentionally because she has a reason for doing what she does: she wants to stay alive. But on the other hand she does not act freely because she is coerced to read the statement and because that statement is no expression of her own beliefs on the matter. Indeed, we should agree that many freedoms have been taken away from the hostage. Furthermore, the conditions she is in clearly make her morally inculpable for reading the statement. However, the libertarian believes, contrary to many compatibilists, that there is more to freedom than the absence of coercion.¹⁰⁵ What free will really is, is an ability to select one among many real possibilities.¹⁰⁶ And even the hostage possesses this ability: she could try to read the statement in a hesitant or sarcastic tone, or she might even resist reading the statement altogether—despite the possibly fatal consequences. Hence the advocate of the simple picture of free will can accept that many freedoms are taken away from those who are politically, socially or physically oppressed. But in her view there also exists a fundamental kind of freedom that is internal to the ability to act and therefore cannot be taken away as long as the ability to act remains intact. That fundamental freedom she calls “free will” and it is exhibited in every intentional action.¹⁰⁷

The thought that free will does not require anything in addition to intentionality—that it is inherent to intentionality itself—is

¹⁰⁵The absence of coercion was already proposed as the the extra *X*-factor that turns action into free action by the early modern compatibilists Hume and Hobbes. The latter, for instance, writes: “Lastly, from the use of the words free will, no liberty can be inferred of the will, desire, or inclination, but the liberty of the man; which consisteth in this, that he finds no stop in doing what he has the will, desire, or inclination to do” [Hobbes 1651/2012, p. 117].

¹⁰⁶What this means is that a freely acting agent has the *de facto* ability to do multiple things, not that she always has an ability to consciously choose between all of her options.

¹⁰⁷This dissertation will thus be defending the view that determinism is not just incompatible with free action, but that it is incompatible with action itself. Helen Steward [2012a] has recently argued for a similar “agency incompatibilism”. For the differences between Steward’s view and mine, see §4.2.1.

not new. In fact we find a surprising ally in Donald Davidson, the founding father of CTA. In his ‘Freedom to Act’ [1973], Davidson tries to analyse ‘*A* is free to do *x*’ as ‘he would do *x* intentionally if he had attitudes that rationalised his doing *x*’ [ibid, p. 79]. In other words, the freedom to do something just lies in the ability—or as he calls it, “causal power”—to do it intentionally. Ultimately Davidson rejects this specific analysis because of the problem of deviant causation, to be discussed in Chapter 3, makes it impossible to empirically identify the causal conditions of intentional action, yet he believes that this is “no obstacle to the view that freedom to act is a causal power of the agent” [ibid, p. 81]. It seems, then, that Davidson himself believed that there is no more to free will than the ability to act on the (causal) basis of one’s own reasons. And on his account being caused to act by one’s own reasons is what intentional action consists in. So he and the simple picture libertarian agree, no extra *X* is needed on top of an account of intentional control. If agents can exert intentional control over their own actions, they possess free will. What Davidson and the libertarian disagree on, is whether such control is compatible with determinism.

2.5 Concluding remarks

This chapter started with an overview of the contemporary free will debate. We saw that the positions in this debate can be framed in terms of the problems of control that beset both libertarian as well as compatibilist accounts of freedom. Although both parties can deal with their respective problems of control if they accept the causal theory of action, this theory also robs them of most good reasons for insisting on determinism or indeterminism. In other words, as soon as the causal theory is accepted, both strong compatibilism and libertarianism collapse into weak compatibilism. That the avowal of CTA involves an implicit commitment

to weak compatibilism is obscured by the contemporary debate which is not focused on understanding agential control, but on proposing and discussing what free will requires in addition to such control. I have argued that the libertarian should not be tempted into defending that indeterminism is such an additional requirement, or that indeterminism merely enhances control. The advocate of the simple picture should argue that control is entirely impossible without indeterminism. Therefore, if the libertarian takes her view that free will requires indeterminism seriously, she should deny CTA and propose an alternative account of control. But such a radical and unshakable commitment to indeterminism is rarely found among contemporary libertarians—and this might explain why it is thought that “[p]erhaps libertarianism is in the last analysis untenable” [Wiggins 1973, p. 270]. At the end of his *Essay on Free Will* Peter van Inwagen, for instance, reluctantly admits:

[I]t is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think, should we become compatibilists [van Inwagen 1986, p. 223]

Similarly, when Robert Kane is asked by John Martin Fisher what he would do if he would ever wake up to a headline reading “Scientist have discovered that determinism is true”, he concedes:

If I do ever read Fischers future headline and it is true, I would give up my libertarian view and perhaps go over to one of these other views. [Fischer, Kane *et al.* 2007, p. 181].

But why should the libertarian commitment to indeterminism be so frail? As a libertarian I believe that we have free will and that free will requires indeterminism. Therefore it is not conceivable to me at all that scientists will ever present me with compelling evidence for determinism. Determinism might be possible, but not in a world that is inhabited by human beings.

Chapter 3

Accidentality in Action I

The Event-Causal Theory of Action

[Davidson] speaks of the possibility of ‘wrong’ or ‘freak’ causal connexions. I say that any recognizable causal connexions would be ‘wrong’, and that he can do no more than postulate a ‘right’ causal connexion in the happy security that none such can be found.

—Elizabeth Anscombe¹⁰⁸

In Chapter 1, I criticised neuroscientists for identifying the intention to perform a movement with neuronal events that occur before the movement. The blame for this mistake is, however not entirely on the neuroscientist. Mainstream philosophy also assumes that the intentionality of an action consists in its having a special causal history. The view that (intentional) action should be distinguished from mere bodily movement in virtue of its event-causes again played an important part in Chapter 2. There I argued that this causal theory of action (CTA) is inherently compatibilist and

¹⁰⁸Anscombe [1974, p. 378].

that it should therefore be rejected by anyone who wants to defend a libertarian position on free will. Be that as it may, CTA (in its many variants) is the principal view in contemporary action theory and has even been described as one of the few achievements of analytic philosophy.¹⁰⁹ Given the wide acceptance of CTA, the result that CTA is inherently compatibilist could easily be seen as a victory for compatibilism. In any case, the result places a heavy burden of proof on those who want to deny compatibilism. Libertarians, it turns out, have to refute the dominant view in the philosophy of action. This chapter is an attempt at such a refutation.

There exist two main worries about CTA. The first is that if agency is a mere matter of event-causation, it becomes hard to see how we can attribute an action to an agent, instead of to her mental events. On CTA the agent merely is the arena in which the mental events do all the work. No role seems to remain for the agent herself. Hence this worry is known as the ‘disappearing agent objection’. Indeed, one can wonder to what extent CTA allows that an action can be *up to* the agent—reminding us of the problem of control described in the last chapter. The second worry is known as the problem of deviant, or wayward causal chains. On CTA a movement is an intentional action if it is caused by a mental state that represents it. But there are cases where we intuitively do not want to call a movement intentional, even though it was caused by a mental state that represented it. In these cases the state caused the action, but not in the right way.

Although these two main problems for CTA are well known, they hardly ever seem to convince anyone to reject CTA altogether.

¹⁰⁹ Julia Tanney [1995] and John Michael McGuire [2007] attribute this statement to Ruth Millikan [1993], but interestingly, it cannot be found in her book. Nevertheless, Millikan has described CTA as such in personal conversation. So, although the origin of the characterisation of causal action theory as one of the few achievements of analytic philosophy remains uncertain, it does seem to describe a widespread opinion about the view.

The disappearing agent objection is often taken to be no more than an incredulous stare, or, at most it is seen as a reason to slightly tweak CTA. The problem of deviant causal chains is taken a bit more seriously. Many philosophers have their token favourite specification of causation in the right way. Those that do not, think the problem of causal deviance only shows that there might not be a general formulation of the causal conditions for intentional action. However, the absence of a general formulation does not shake their belief that we know perfectly well what a right causal chain is in the individual case.¹¹⁰ In this chapter I will argue that these problems should not be so easily dismissed. Contrary to general belief, deviant causal chains are no mere oddity, they are symptoms of a deeper, much more serious problem for CTA. As I will argue, CTA lacks the resources to explain how someone can perform an action *on the grounds* of her reasons for it, i.e., how her actions are caused *in virtue of* the content of her reasons. Therefore, even when seemingly no deviance occurs an action is only *accidentally* in accord with the mental state that causes it. That the agent moves in accordance with her thought turns out to be a mere matter of luck, and hence CTA fails as an account of agential control.

I will proceed as follows. First I will sketch CTA and its background (§3.1). I continue by discussing the disappearing agent objection, and show that it is inconclusive on its own (§3.2). The discussion of the problem of causal deviance and the many proposed solutions will make up the majority of the chapter (§3.3). Finally I will offer a diagnosis of the deep problem for CTA (§3.4), and show what implications this problem has for our understanding of free will (§3.5).

¹¹⁰Late in his life Davidson himself adopted such a dismissive attitude. See §3.4.

3.1 The causal theory of action

Although the causal theory of action nowadays is the considered to be the “standard story” [Smith 2004, p. 165], when Donald Davidson first proposed it he considered himself to be swimming against a “very strong neo-Wittgensteinian current of small red books” [Davidson 2001, p. 261].¹¹¹ These neo-Wittgensteinians argued that there are two very different kinds of enquiry into why something happens. One can answer the question “Why did the window break?” by citing a *cause*: “It was struck with an emergency hammer.” But the question can also be interpreted as asking for an *aim* instead of a cause. Then a possible answer could be: “because I wanted to get out of the burning car”. Hence, the same event can have two different explanations: one causal, the other teleological. Causal explanation reveals why an event occurred by pointing to another event that brought it about. Teleological explanation rationalises the event: it places the event in a context that shows what reasons there were for it to happen. According to the neo-Wittgensteinians, action explanation is strictly teleological. If we want to know why someone φ -ed we are after the agent’s reasons or motivations for φ -ing. This investigation, they think, is wholly independent of the enquiry into its causes—the latter is perhaps only of interest for the (neuro)-biologist. Davidson took issue with this separation of causation and rationalisation.

In his ‘Actions, Reasons, and Causes’ [1963], Davidson argues for the thesis that reasons *are* causes: rationalisation is a species of ordinary causal explanation. His argument is clear and convincing. We have reasons to do many things (sometimes incompatible things), yet we do not act on every reason we have. Having a reason to φ is simply not sufficient to explain φ -ing, because

¹¹¹At the time the books published in the Routledge & Kegan Paul series *Studies in Philosophical Psychology* edited by R.F. Holland were indeed small and red. Included in the series were Anthony Kenny’s *Action, Emotion and Will* [1963], and A.I. Melden’s *Free Action* [1967].

even one who has reason to φ might not φ . To understand why an agent performed an action, we do not only need to know that the agent had a reason to perform the act, but also that she acted *for* that reason. Teleological explanation shows why performing a certain action would be rational, but only causality, Davidson argues, can account for an action's actual occurrence. Consider Charlie who has a very annoying and obnoxious doctor and therefore has a reason to kick her. As a matter of fact she does kick her doctor but only because the doctor was testing her knee reflex while being too close to Charlie's lower leg. In this case Charlie did of course not intentionally kick the doctor even though she had a reason to do so. It was a *mere accident* that Charlie's movements were in accord with her reason. What is required for the kicking to be intentional instead of accidental, is that Charlie kicked the doctor *because* she was obnoxious and annoying.

But is this 'because' that signifies the relation between an action and the reason for which it was performed, the same 'because' as that of ordinary causal explanation? After all, a physical event like the striking of the window with a hammer seems quite different from a reason. Davidson argues that the difference between a reason and a physical cause, is not that the one is an event where the other is something else, but that the one is physical and the other mental. Reasons are belief-desire pairs: the reason for breaking the window consists in the desire to get out of the burning car and the belief that breaking the window can help one get out.¹¹² So far so good, but how can these *mental* events of believing and desiring bring about a *physical* movement? Davidson famously believed that this problem of mental causation can be solved if you consider that every mental event can also be described in physical

¹¹²Often, beliefs and desires are referred to as mental *states* and mental *events* are identified with the onsets of mental states. This difference does not matter much to my purposes, since both are mentioned to play the role of mental cause by different causal action theorists. Therefore I will use the two terms interchangeably.

terms (for instance as a happening or a complex of happenings in the brain).¹¹³ Hence, reasons can cause movement just like one billiard ball can cause another to move.

Although Davidson thinks that rationalisation is a species of causal explanation, there must be more to rationalisation than mere causation. Not every movement that is caused by a reason is rationalised by it. I might desire Kung Pao chicken and believe that there is some left in the fridge, which causes my mouth to water. Now the movements in my salivary glands are not movements I intend to make. Desiring Kung Pao chicken does not give me a reason to water my mouth. According to Davidson, what is needed in addition to the causal relation between mental state and movement, is that the mental state represents the movement as valuable or desirable to the agent. When mental states about the Chinese dish cause me to get up and walk to the fridge, that movement is intentional because I take going to the fridge to be conducive to attaining what I desire.

Davidson concludes that the explanation of action is neither solely teleological nor solely causal:

The action on the one hand, and the belief-desire pair which give the reason on the other, must be related in two very different ways to yield an explanation. First, there must be a logical relation. Beliefs and desires have a content, and these contents must be such as to imply that there is something valuable or desirable about the action [...] Second, the reasons an agent has for acting must, if they are to explain the action, be the reasons on which he acted; the reasons must have played a causal role in the occurrence of the action. [Davidson 1982, p. 141]

We can *fully* explain an action by means of these two separate relations. What it is for a bodily movement to be intentional, is that it is caused by a reason that represents it. There is nothing more to it.

¹¹³See, Davidson [1970].

We can now give a general formulation of the causal theory of action:

(CTA) A bodily movement is intentional if and only if it is *caused* by a thought that *represents* it.¹¹⁴

This account of action is a reductive account. It understands the complex concept of intentional action in terms two more fundamental and independently intelligible concepts: representation and causation. An intentional and a non-intentional movement are not *intrinsically* different on Davidson's view. Where they differ is in their extrinsic relations to prior events. Hence we can think of CTA as a "decompositional analysis" of intentional action.¹¹⁵

Since its original publication, there has been much discussion of Davidson's account. A main point of contention is Davidson's idea that reasons are belief-desire pairs. David Velleman [1989], for instance, thinks that it is the agent's knowledge of her action, instead of her belief-desire pair, that efficiently causes it. Michael Bratman [1987, 2014] argues that we must recognise intentions themselves as a *sui generis* kind of mental state responsible for the causation of action.¹¹⁶ Furthermore, Davidson's views on the relation between mind and body and his views on causation are widely challenged.¹¹⁷ But none of this has overthrown the core idea of CTA that an action is intentional just if it is caused by a thought that represents it. As Hans-Johann Glock states:

There is no doubt that Davidson succeeded in stemming the "neo- Wittgensteinian" tide of "small red books". His article

¹¹⁴Often CTA is defined as the position that intentional action is caused "in the right way" by the thought that represents it. I will come back to this extensively in §3.3.

¹¹⁵The phrase "decompositional analysis" is due to Doug Lavin [2013].

¹¹⁶Davidson later [1978] also accepts that intentions are themselves mental states that play a role in the causation of action. Noteworthy proponents of causal action theory that do uphold the belief-desire model are Goldman [1979] and Dretske [1989].

¹¹⁷See, e.g., Fodor [1992], Searle [2001] and Kim [2007].

reinstated causalism not merely as the majority view, but as something that is widely regarded as a truism. [Glock 2014, p. 17]

3.2 The disappearing agent

In his ‘What happens when someone acts?’ David Velleman introduces the disappearing agent objection to CTA:

In [the standard] story, reasons cause an intention, and an intention causes bodily movements, but nobody—that is, no person—*does* anything. Psychological and physiological events take place inside a person, but the person serves merely as the arena for these events: he takes no active part. [1992b, p. 461]

But can’t proponents of CTA have a quick and ready response to this objection? Their theory was supposed to offer a reductionist account of intentional agency. They believe that the story about the psychological and physiological events *is* the story about what it is for an agent to act. Agency just *consists* in the occurrence of appropriately related physical and mental events. As Velleman admits, demanding a role for the agent in the event-causal story seems like demanding that a cake needs to be listed in its own recipe. Nevertheless, he deems that response inadequate. His reason for this, however, is not that the agent herself needs to be mentioned in the standard story, it is, rather, that the events that are included in the story cannot add up to a person’s activity. In the original version of CTA, an action is caused by a desire for a particular end in conjunction with a belief that the action is conducive to the end. Velleman argues that an action is only directly caused by a belief-desire pair in the exceptional case: an addict may desire a drug and believe that injecting the substance in the syringe satisfies that desire, and before she knows it, the needle is in her arm. But not every act is like that of the addict. In general, agents often reflect on their beliefs and desires before they

act. Reflection, according to Velleman, is the primary role played by the agent when she acts, and standard CTA fails to account for this role. In Velleman's view this failure is no argument to abandon CTA. It merely provides a reason to amend it. The role of the agent can be accounted for in event-causal terms when we add a further desire to the story: the desire to act in accordance with reason. It is by means of this desire that the agent can throw her causal weight behind those reasons she believes to be best.¹¹⁸ It thus turns out that Velleman is not objecting to the reduction of agency to psychological and physiological events that take place in the agent. He is merely objecting to a reduction to certain kinds of event. In this regard Velleman's criticism of CTA is like that of Bratman [1987] who also believed that agency cannot fully be reduced to belief-desire and argued for the inclusion of intentions. Both Velleman and Bratman do not argue against the decompositional project. They just argue against a particular decomposition.

Nevertheless the opponents of CTA, few as they are, seem happy to grasp any straw that can help them in their quest, and some have thus taken Velleman's rhetorical description of the person as a mere arena for the events that make up her agency, to be a serious objection to any version of CTA.¹¹⁹ Their rejection of the event-causal story is usually coupled with a commitment to an agent-causal story:

Human agents are not merely things within which things happen, and they clearly do play a role in the arena within which their actions are found. For an event of someone's doing something is typically an event of her bringing something about. [Hornsby 2004b, p. 177]

¹¹⁸Here is not the place to discuss whether this amendment successfully captures the agent's role. Franklin thinks it does not, but he also does not see that as a reason to abandon CTA, instead he offers a slightly different mental state that, according to him, *can* do the job. See Franklin [2014].

¹¹⁹See Hornsby [2004a,b] and Steward [2012a].

I will have much more to say about agent causation in Chapter 4 and I will end up defending a view that can be considered agent-causal in chapters 5 and 6. In the present context, however, we only need to note that reiterating a commitment to agent causation *cannot* in itself be an argument against CTA. That theory after all is an attempt to understand agent-causation in non-agential terms. As Bishop clearly states:

[CTA maintains] that my raising my arm consists in my arm going up as a causal consequence of, say, my intention to raise my arm. On this theory, agent-causation is (in a certain sense) reducible to ‘ordinary’ causation by mental events, and so it fits unproblematically into the ontology of scientific naturalism. [Bishop 1989, p. 2]

In the face of this, the disappearing agent objection indeed just comes down to the complaint that the cake itself does not figure in its recipe. Jennifer Hornsby might well be right that agency can never be found in a mere causal succession of events, but the disappearing agent objection does not help to establish this. Pointing out the simple fact that CTA is reductive, cannot in itself be an argument against reduction. The disappearing agent objection does offer a nice rhetorical device to unearth the reductionistic commitments of the causal action theorist, but as an argument against CTA it is hopelessly question begging. In order to bring down CTA, we need a further argument that shows why agency cannot be successfully reduced. The remainder of this chapter is devoted to constructing such an argument.

Before we move on, however, we should stop to note a similarity between the disappearing agent objection and the luck problem we came across in §2.1. That problem challenged the extent to which an action in an indeterministic universe can be *up to* the agent, or under the agent’s *control*. The disappearing agent objection can also be viewed as a challenge to control. If the agent is only the

arena in which the events that make up her agency occur, then how can her actions be up to her? Derk Pereboom [2004, 2007, 2015] notices the similarity and argues that the luck problem and the disappearing agent objection are part of the same family. The latter, he believes, reveals “the deepest problem for event-causal libertarianism” [Pereboom 2015, p. 7]:

On an event-causal picture the relevant causal conditions antecedent to a decision—agent-involving events, or alternatively, states of the agent—would leave it open whether the decision will occur, and the agent has no further causal role in determining whether it does. [Pereboom 2007, p. 102]

And hence, Pereboom concludes, an agent has no control over her decisions, they can only be up to luck.¹²⁰

Let us assess whether this luck argument in the form of a disappearing agent objection should impress the event-causal libertarian. I think it should not. Remember that event-causal libertarians endorse CTA, a reductive account of agency. On that view an agent’s decision *consists* in the causal battle between her antecedent mental states. Complaining that the agent has no further causal role in determining which set of reasons wins the battle, again comes down to complaining that the cake is not listed in its own recipe. CTA fully reduces the agent’s causal role to causation by mental events. If CTA is the correct account of causal control, then Pereboom’s objection demands something that cannot and should not be had.¹²¹ The disappearing agent objection to event-causal libertarianism only sticks, it turns out, if CTA does not provide the correct account of agential control. From this, we can

¹²⁰Timothy O’Connor [2009a] and Meghan Griffith [2010] who both endorse agent-causal libertarianism agree that this objection is decisive against the event-causal libertarian.

¹²¹In §2.2 where we discussed Franklin’s reply to the luck objection, we saw that Schlosser [2014] also failed to fully appreciate the reductive nature of event-causal libertarianism in complaining that agents on Franklin’s account lack the power to select which alternative action they will pursue.

draw a conclusion that has not yet been recognised in philosophical literature: if event-causal libertarianism fails to secure control, then it fails not because of its libertarianism (i.e., its endorsement of indeterminism), but because of its event-causalism. It follows that if CTA is incorrect, then non-libertarian (i.e. compatibilist) accounts of control in terms of CTA also fall prey to this version of the problem of luck. This is a surprising conclusion, given the fact that the luck problem is traditionally conceived as an issue for libertarians only. I will come back to this observation at the end of this chapter (§3.5).

3.3 Deviant causal chains

The problem of causal deviance, or waywardness, is that some bodily movements are clearly unintentional, despite the fact that they are caused by mental states that represent them. Davidson himself gave the most famous example of such a deviant scenario:

A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. The belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. [Davidson 1973, p. 79]¹²²

The climber's movement is caused by mental states that represent it, but it nevertheless is no intentional movement. Therefore, the original formulation of CTA presented in section 3.1 cannot be correct. As we saw there, Davidson insisted on a causal relation between reason and action because sometimes one can accidentally

¹²²This is an example of antecedent or primary deviance: deviance in between the mental state and the movement. There also are cases of secondary or consequential deviance: deviance in between the movement and its consequences.

move in accordance with one's reasons—as was the case with Bo who kicked her doctor. The example of the climber shows that a causal relation between thought and movement is insufficient to rule out accidentality. In that scenario the thought caused the movement but in an accidental way. To rule out such accidental causation, Davidson [1978, p. 87] contends that CTA should be amended as follows:

(CTA) A bodily movement is intentional if and only if it is caused *in the right way* by a thought that represents it.

Of course, the million dollar question now becomes: what *is* the right way in which the mental state should cause the movement it represents?¹²³ *Prima facie*, there seems to be a straightforward answer to this question: A mental event *e* causes a movement *m* in the right way if and only if the causal process from *e* to *m* is such that it realises agential control. But although the causal theorist is indeed after that causal process that realises control, this answer is not available to her. CTA aims at an analysis of intentional action in non-agential terms, hence the proponent of CTA must find a manner to spell out causation in the right way without making reference to agency or agential control, i.e., it has to reduce control to something that is altogether non-agential.

Much more ink has been spilled on the problem of causal deviance than on the problem of the disappearing agent. No causal action theorist, however, sees causal deviance as destructive to her position. Attitudes towards the problem generally come in two variants. Certain philosophers take up the challenge to specify causation in the right way and believe that they are able to do so. Others believe that philosophy cannot come up with a general solution to the problem, but they believe that the problem is only a marginal one: we are perfectly able to recognise a correct causal

¹²³Davidson admits that his account as it stands is an “incomplete and unsatisfactory account of acting with an intention” [Davidson 1978, p. 87].

chain when we see one. In the next section (§3.4) I will argue against such a dismissal by claiming that deviant scenarios are no mere oddity, but the symptoms of a deeper problem for CTA. This section will concentrate on the proposed specifications of the right way and explain why each is unsuccessful.¹²⁴

3.3.1 Causation in the right way

Let us start by considering two strategies for understanding causation in the right way: the causal immediacy strategy, and the sustaining causation strategy.

The causal immediacy strategy. Causal deviance occurs because something intervenes between the mental state and the action. In the climber case, for instance, the agent's desire and belief do not directly bring about her movement: they first bring about a nervousness and this nervousness causes the loss of grip. The idea behind the causal immediacy strategy is to leave no room for possible interventions. Proponents of this strategy require that the mental state *proximately* or directly causes the movement it represents.¹²⁵ Apart from worries one might have about the very conceptual coherence of proximate causation, it is particularly dubious to employ the notion in the context of the causation of action. Any bodily movement is at least preceded by events in the agent's muscles, in the nerves that activate these muscles and in her motor cortex. Hence no intentional movement can be proximately caused by a mental state.

¹²⁴In what follows I have tried to chop up the discussion on the extra conditions that are supposed to rule out deviancy as neatly as possible. Many causal action theorists, however, do not stick to one strategy and use hybrid approaches to deal with different cases. John Bishop, for example, combines the sensitivity and sustaining causation strategies, see Bishop [1989, pp. 167-71].

¹²⁵See, e.g., Brand [1989], Mele [1992]. A variant of this strategy is to require that the mental cause and movement should occur simultaneously, see Searle [1983].

Even when there is no deviance, events occur in between the thought and movement, but these events do not disrupt intentionality: they are events of the right kind. But trying to determine which events are of the right kind, of course, just is trying to specify causation in the right way. To avoid the re-emergence of the problem the causal immediacy strategy set out to solve, it could be argued that intentional action includes the entire neurophysiological chain that leads to the bodily behaviour, the beginning of which is proximately caused by a mental state.¹²⁶ But not only is it philosophically dubious to think of the neurophysiological events as part of the action¹²⁷, it also merely displaces the problem of causal deviance. Although no deviance can occur in between the mental state and the start of the neurophysiological chain, how can we rule out the possibility of deviance *within* that chain? An evil demon—or its contemporary equivalent: the clever neuroscientist—could concoct a device that detects when the chain has started, paralyses the muscles, but then makes sure the represented movement happens anyway by pushing or pulling the body.¹²⁸ We must conclude that proximate causation does not equal causation in the right way.

The sustaining causation strategy. Harry Frankfurt [1978] has famously criticised Davidson's account of agency on the ground that it requires nothing of the agent after the act is initiated. This, Frankfurt argues, does not cast the agent in its proper role. When we act, we do not just initiate our movements, but *guide* them all

¹²⁶Cf. Mele [1992, p. 202]

¹²⁷One argument is that we are typically conscious of our intentional movements, whereas we are normally unconscious of the neuronal and physiological events that precede them. Much more on the relation between intentionality and consciousness will follow in Chapter 5.

¹²⁸We could say that antecedental waywardness (deviance in between the thought and action) is only ruled out at the expense of increasing the space for consequential waywardness (deviance in between the start of the action and its consequences).

the way through to their conclusion. Might it be Davidson's focus on the commencement of the movement that opens his theory up to deviance? In the climber example, although the climber's thoughts cause his grip to loosen, he is not guiding this movement as it happens. Hence, many philosophers propose that CTA should not analyse intentional action as merely brought about by a thought that represents it, the thought should also regulate, control and sustain the performance of the movement.¹²⁹ When we act, the idea is, we do not initiate our actions and wait for our movements to conclude, we constantly correct the path of our movement to ensure that we attain our ends.

If the notion of guidance is of help to CTA's reductive agenda, there must be a way to spell out guidance in non-agential, purely causal terms.¹³⁰ One idea is that the mental state that produces the movement should not just kick off the movement but should continue to be actively present. But this is insufficient, for a mental state could generate and sustain a nervousness which then (deviantly) causes a bodily movement.¹³¹ A less crude idea is proposed by Joseph Raz:

The element of guidance can be understood by analogy to a negative feedback mechanism: we, automatically and normally without being conscious of the fact, monitor the performance of the intentional action such that if it deviates from the course we implicitly take to lead to its successful completion we correct the performance, bringing it back to the correct path. [Raz 2009, p. 199]

Being produced and sustained by a negative feedback mechanism is by itself not sufficient for a movement to count as intentional. The dilation of the pupils in response to a change in light,

¹²⁹See, e.g., Thalberg [1984], G. Wilson [1985], Adams and Mele [1989], Velleman [2000], Setiya [2007] and Raz [2009].

¹³⁰There are attempts to spell out guidance in counterfactual terms. I will discuss this proposal when I consider the sensitivity strategy below in §3.3.3.

¹³¹Mele [1992, p. 203] recognises this worry.

for instance, is no intentional movement. What the causal action theorist needs, is that *the intention itself* causally sustains the behaviour it brings about by way of negative feedback loops. However, as Erasmus Mayr [2011] shows, the obtaining of such feedback processes is neither necessary nor sufficient for agency. It is not necessary because there are actions, such as the finger movements of an accomplished clarinettist, that are too fast to allow for feedback.¹³² But, more importantly, it is insufficient because there can be causal deviance within the feedback loops. The clever neuroscientist we encountered above again spoils the party:

[The neuroscientist] now interferes in the stage between the motor command and the motion, causing the latter in accordance with what he ‘reads off’ from the former, and then allows the feedback loop to go back to the agent’s central nervous system where the information continues to be processed. [ibid, p. 128]

The sustaining causation strategy suggests that causal deviance can be ruled out by turning action into a continuous causal process rather than a singular causal chain. As we now see, this can be thwarted simply by constructing a deviant scenario where the neuroscientist’s interventions are also continuous and adaptive.

3.3.2 Representation in the right way

The two strategies discussed above try to battle causal deviance with extra requirements on the causal process that leads from intention to action. The first requires intention to be the proximate cause of action, the second requires a continuous feedback process between movement and intention. As we have seen, neither strategy succeeds to rule out the possibility of causal deviance. The strategy I will discuss next focuses on the relation of repres-

¹³²For more examples and an elaborate discussion of actions without feedback, see Mayr [2011, pp. 128–130].

entation. The strategy does not attempt to rule out deviance by specifying causation in the right way, but by singling out the right kind of representation.

Intending to act according to plan. The examples of causal deviance seem to show that it is possible that an agent φ 's while she has the intention to φ , without the agent's φ 'ing being intentional. Gilbert Harman [1976] denies that the agent in the deviant case does what she intended to do. According to him an intention does not just represent an action, it also includes a "plan specifying how that intention will lead her to do what she intends to do" [p. 161]. In cases of deviant causation one does not do what one intends to do, because one does it in a way that was unintended.¹³³ Harman demonstrates how his suggestion works with the help of the following case: a nephew wants to kill his uncle by shooting him, but instead the nephew kills his uncle by accidentally running him over when he is on the way to his uncle's house to execute the murderous plan.¹³⁴ In this case the nephew intended to kill the uncle, and this intention caused him to kill the uncle (by causing him to drive over to his uncle's house), but he did not kill his uncle intentionally. CTA would wrongly deliver the verdict that the nephew did intentionally kill his uncle, since the killing was both caused and represented by the nephew's thought. But Harman's modified account delivers the right verdict. On his account the agent's thought not only represents the killing, but also represents the way in which it is to be achieved (by shooting), wherefore Harman can deny that when the nephew runs over the uncle, the killing was caused by a thought that also represented it.

¹³³Several other authors e.g., Bratman [1987], Thalberg [1984] and Brand [1989], also believe that the causal analysis can be fortified if you demand that intentions include action plans. These authors, however, do not believe that including the action plan in the intention is sufficient to deal with every case of causal deviance.

¹³⁴This example is due to Roderick Chisholm [1966].

The problem with Harman's suggestion, however, is that we can quite easily come up with more devious causal scenarios that do conform to the agent's action plan. Suppose the nephew makes it to his uncle's house without running him over, and upon spotting the uncle in his garden, fires a shot. The shot however misses the uncle by a few inches, nevertheless the uncle meets his demise because the shot caused a nearby herd of wild boars to stampede and trample him to death.¹³⁵ Now Harman might of course reply that in this case also, the nephew did not kill his uncle in the way he intended to. He intended to kill him by having the bullet penetrate one of the uncles vital organs. There are at least two problems with this reply. First it seems easy to come up with an even more devious case that conforms to the new action plan: a nearby hunter (who was likely after the herd of boars) is out of ammo and finds the nephew's bullet after it missed the uncle, the hunter loads this bullet into his gun, takes aim at a boar, but right when he is pulling the trigger the hunter slips, which causes the bullet to fly off in a wayward direction and, as you can guess, ends up penetrating one of the uncle's vital organs. Second, it seems that the nephew's plan has become too specific. Would the nephew really care whether he kills his uncle by having his bullet penetrate a vital organ, or by actually having it destroy a major blood vessel? The action plans we have, cannot foresee every possible deviancy and rule it out. To illustrate: deviance can also occur in the neurophysiological chain leading up to the movement (as the clever neuroscientist from above shows), and we do not have enough knowledge of, or control over neurophysiology to plan out that part of the chain. As Geert Keil [2007] argues, an agent can only retrospectively identify whether the actual causal chain from her mental state to her action conforms to her action plan. That deviancy can only be established after the facts, is

¹³⁵This example is due to Daniel Bennett, cf. Davidson [1973].

“bad news” [p. 84] for CTA, since that theory contends that the mark of the intentional is that a mental state prior to the action both rationalises and causes it.

3.3.3 Causation because of representation

Above we have seen attempts to rule out causal deviancy by either exclusively restricting the causal relation between thought and behaviour, or by exclusively specifying the relation of representation. Advocates of the strategies to be discussed next believe that CTA allows for deviance not because something is wrong with the causal relation or the representational relation themselves, but because the caused behaviour is not caused in response to the represented content.

The sensitivity strategy. Consider an agent who is caused to buy some cat food by her intention to buy cat food. Had the content of her intention been different, e.g., had she intended to buy cat litter, we would expect her action to be different as well: she would buy litter instead of food. Now in the case of the climber the behaviour does not seem to depend on the content of the climber’s intention. Had the climber desired to jump down himself in order to be done with the arduous ordeal of climbing, then this thought might equally well have brought about a nervousness that caused him to let go of the rope. The difference between the normal causal chain of buying cat food and deviant causal chain of the climber case, is that the behaviour in the first case is *sensitive* to the content of the intention, where the behaviour of the climber is not. Hence some philosophers advise to add a sensitivity requirement to CTA.¹³⁶

The sensitivity condition of course has to be spelled out in reductive non-agential terms. This is either done in terms of explan-

¹³⁶See, e.g., Peacocke [1979], Bishop [1989].

atory connections between intention and behaviour, or in terms of a counterfactual dependency between the two. Both proposals suffer from the same problem, so for the sake of brevity I will only focus on the latter account of sensitivity.¹³⁷ The counterfactual account of sensitivity states that a bit of behaviour is sensitive to the content of the intention, if and only if, had the content been different, the resulting behaviour would have been correspondingly different.¹³⁸

Although a sensitivity condition rules out the climber case, we can easily construct a scenario where there is deviance despite sensitivity. In fact we already came across such a case when we discussed the neuroscientist who reads off an agent's intention, paralyzes her muscles and then makes an external device move the body in accordance with the read-off intention.¹³⁹ Because non-intentional movements can also be sensitive due to accidental circumstances, we can conclude that sensitivity by itself does not save CTA.

Christopher Peacocke, however, reacts to such examples by stipulating that the “chain from intention to movement must not run through the intentions of another person.” [1979, p. 88] As Mayr [2011, p. 118] points out this move is very *ad hoc*, and the need for such an additional condition “makes it doubtful that sensitivity really provides the key to understanding causal deviance.”

¹³⁷See Mayr [2011, §5.4] for a parallel discussion of the former way of spelling out sensitivity.

¹³⁸Sometimes, for example in Adams and Mele [1989], the notion of guidance we came across in §3.3.1 is defined counterfactually: the agent guides her movements if she would have intervened had her movements gone astray. This approach combines intuitions about guidance and sensitivity. However, it is not better off than the sustaining causation approach discussed above, since, as we will presently see, both the guidance and the sensitivity approach fall prey to the same counterexample.

¹³⁹Alternatively we can add a neuroscientist or demon to the climber case who makes sure that the agent acts differently if the content of his intention is different so that the counterfactual requirement is satisfied. In that scenario, although the climber's behaviour is sensitive to counterfactual differences, this does not make his actual loosening of his grip any less deviant.

Furthermore, there are genuine cases of intentional action where the causal chain from intention to movement does run through another person's intentions. This often happens in cases of cooperative action: I might intend to tune my trombone, but I am only able to do so if another person first plays a note of 442hz.¹⁴⁰ All this shows, I believe, that the sensitivity strategy gets the order of explanation wrong. When an action is intentional, it is performed *for* a reason. It is because of this, that the action would have been different had the reason been different. The fact that a movement is intentional explains why it is sensitive to content. The sensitivity strategy errs because it tries to explain what acting for a reason is, in terms of sensitivity. But as we have seen a bit of behaviour can be accidentally sensitive and therefore sensitivity does not guarantee intentionality.

Causation in virtue of content. In deviant causal chains cases, it is just a fluke that the antecedent mental state both rationalises and causes the action. In case of a genuine intentional action, the fact that the thought causes the action should not be wholly independent of the fact that the action is rationalised by the thought. The sensitivity requirement failed to rule out such accidentality because a bit of behaviour's sensitivity to the content of a mental state can itself be accidental. Nevertheless, the thought that causal chains are deviant when causation is independent of content seems true. What sets an intentional action apart, is that it is brought about by a mental state *in virtue of* the content of that mental state.

The reason why the causal pathway in the climber example is deviant seems to be that the movement is merely caused by the reason states—the movement is not a response to the reasons *qua* reasons [...]. The reason state causes and rationalizes the

¹⁴⁰This is an example of what Michael Bratman [2014] calls 'modest sociality'.

action, but the reason state's rationalizing the action seems to be irrelevant to its causing it. [Schlosser 2007, p. 191]

Recently, a growing contingent of philosophers (e.g., Arpaly [2006], Wedgwood [2006] and Schlosser [2007]) have argued that 'causation in the right way' just comes down to causation in virtue of content. But how are we to understand what causation in virtue of content is? What is it for a reason not to merely cause a movement, but to cause it *qua* reason? Interestingly the philosophers who defend this strategy do not think that a special answer to this question is needed. Normy Arpaly [2006, p. 47] considers a counterfactual answer—something is caused in virtue of content if it would not have come about if the content had been different—but quickly (and rightly) dismisses it because this answer brings her strategy dangerously close to the sensitivity strategy and opens it up to the same counterexamples. But instead of further attempts at clarifying what causation in virtue of content is, she retreats to a misery loves company defence:

[I]t is not clear exactly what it means for something to cause something "by virtue" of a particular fact about it. But again, this problem is not particular to actions or to cases involving mental states: it is equally hard to explain what it means for a rock to break a window by virtue of its size and weight. [ibid, p. 71].

Ralph Wedgwood offers a similar sentiment:

This formulation—'causation in virtue of rationalization'—has been used by other philosophers in discussing the problem of deviant causal chains.^[141] None of these other philosophers, however, regard this formulation as itself giving the solution to the problem. Instead, they regard this formulation as simply another statement of the problem itself. This is because they think that the phrase 'in virtue of' must have some different meaning

¹⁴¹Louise Antony [1989, p. 168] and Bill Brewer [1995, p. 238,247].

here from its ordinary use in causal explanations—such as in the claim that the impact of the planes caused the towers to collapse in virtue of the temperature of the ensuing fuel explosion, and not in virtue of the force of the impact. My proposal is that in the formulation ‘causation in virtue of rationalization’, we should take the phrase ‘in virtue of’ in exactly the same sense as in those ordinary causal explanations. [Wedgwood 2006, p. 670]

Is this attitude justified? In the cases of ordinary causal explanation we are able to give a story of what it is for an event to cause another event in virtue of one property rather than another. Upon investigating the collapsed tower, for instance, we can find out that the destruction was primarily caused by the expansion and melting of the metal frame, and not by, say, the bending or breaking of the frame. As science tells us, metal reacts to extreme heat by expanding and melting and to heavy impact by bending or breaking, hence we can conjecture that the tower collapsed in virtue of the heat of the explosion and not in virtue of its impact. By looking at the mechanisms involved in the collapse we can understand which features of the cause-event were crucial for the effect to occur. But what mechanisms are involved in the causation by a reason, *qua* reason? And how are these mechanism different from those involved in the causation by a reason *qua*, say, neuronal-event?¹⁴²

Suppose, if only for the sake of argument, that these questions can be answered, then it still has to be seen whether the answers can be given in non-agential terms. Remember that CTA is an attempt at the decomposition of agency into the more basic concepts of causation and representation. The defenders of the current strategy are right, I believe, that agency is characterised by causation in virtue of content, but why do they think that CTA

¹⁴²Some of the strategies discussed above can be viewed as attempts to answer these questions. Proponents of the sustaining causation strategy, for instance, argue that mechanisms involved in intentional action are different from non-intentional mechanisms due to the feedback loops they contain.

can give us an account of that? When they claim that a third condition should be added to CTA—the mental state should not just cause and represent the movement, it should also cause it in virtue of the representational content—Wedgwood etc. are reintroducing intentionality in the analysis. CTA was supposed to be a reductive analysis of agency, that is: an analysis of causation by reasons *qua* reasons, and hence causation by reasons in virtue of content cannot itself be a part of the analysis. By itself the addition of the ‘in virtue of content requirement’ does no more work than claiming that the mental state should not only cause and represent the movement, but should also cause it ‘in the right way’. Therefore, contra Wedgwood, Anthony and Brewer are correct when they regard the formulation “causation in virtue of rationalisation” as a mere restatement of the problem itself. Moreover, they might not think this because they believe that “the phrase ‘in virtue of rationalization’ must have some different meaning” when used in the context of the mental, but because they think it needs to have *exactly the same* meaning. It is only that we have no idea how we are to understand what it means for a reason to bring something about in virtue of its content.

3.3.4 Miscellaneous strategies

In the next section I will argue that CTA is doomed to failure because it is in fact impossible to give a reductive story of causation in virtue of content. But before we go there, I will shortly consider some more exotic, out of the box, attempts to overcome causal deviance.

Berent Enç’s evolutionary strategy. In cases of deviance, the intention causes the movement it is supposed to cause, but it does not cause the movement in the way it is supposed to cause it. The normal function of our intentions is to cause intentional ac-

tions, hence when they waywardly cause a movement that is not intentional, this constitutes a *malfunctioning*. Therefore it seems that we are able to rule out causal deviance if we can give the conditions for the well-functioning of the action generating mechanism. At first sight, these observations might seem unhelpful to the reductive ambitions of the causal action theorist because the notion of well-functioning itself seems hopelessly teleological. However, there might be perfectly naturalistic analyses of proper functioning.¹⁴³ For biological organisms, a historical account of the fitness enhancing features developed and retained in a population reveals the biological function of these features. We can say that flowers emit scent to attract insects, because scent-emitting flowers have been more successful at cross-pollinating and hence at reproduction than their non scent-emitting counterparts (or predecessors). Berent Enç attempts to employ such a historical notion of proper functioning in the context of intentional action. On his view, “the function of the intention-generating system is to produce actions that match what is intended” [Enç 2004, p. 159]. But how are we to rule out that the intention-generating system produces movements that only accidentally match the intention? To answer this question, Enç argues, we should also look at evolution. Non-necessary or accidental steps in a causal process are filtered out by natural selection. A feline species that cleans her fur after every bite of food she takes has, for obvious reasons, a lower chance at survival than an otherwise equivalent species that only cleans up after the entire meal. Similarly, we expect that in non-wayward operations of the intention generating system, every intermediate stage in the chain from intention to movement is explainable by the fact that it was selected for its capacity to contribute to the production of the output which the system has the function of producing. When we return to the climber case, we

¹⁴³Ruth Millikan’s [1984, 1989] account of proper functioning is a prime example.

see that the fact that the intention causes an intermediate stage of nervousness has nothing to do with the fact that the nervousness results in letting go of the rope, and hence this case must be ruled out as deviant. Enç arrives at the following statement of CTA:

(CTA) The behavioural output of an organism is an intentional action (A) if it is caused in the way it is supposed to be caused by an intention to do A. [Enç 2004, p. 157]

But even if we accept both the evolutionary analysis of proper functioning, and the idea that natural selection filters out all the accidental steps in causal chains, a problem remains for Enç's account. On his view, any behavioural output that is caused by an intention in the way it is supposed to be caused is an intentional action. However, as Mayr [2011, pp. 132–133] points out, the proper biological function of intentions is not limited to action generation. A proper function of the intention to eat, for instance, is to cause an increase in the production of saliva required for healthy digestion. Since the saliva production is a behavioural output caused by an intention to eat in the way it is supposed to be caused by an intention to eat, saliva production—or any other physiological change required for the performance of an intentional action—comes out as intentional on Enç's account. Obviously, when we eat, we do not also intentionally produce saliva and therefore we have to dismiss Enç's version of CTA. As Mayr states: “Enç's proposed analysis of action is defeated by the fact that the biological role of intention-states in the production of physiological changes is not restricted to the production of actional behaviour” [ibid, p. 133].

Even so, it seems that if we tweak Enç's account a little it can escape Mayr's criticism. In addition to the requirement that behavioural output should be caused by an intention in the way it is supposed to be caused, we could demand that the intention should also *represent* the behavioural output—as in the standard version

of CTA. The intention to eat does represent eating, but it does not represent, say, saliva production. Therefore, this modified account does not wrongly take every physiological change that is supposed to be caused by an intention to be an intentional action. However, we might still wonder whether this modified account really rules out all accidentality. Nothing in the account seems to guarantee that it is impossible for an intention to only accidentally causes a movement it represents in the way it is biologically supposed to cause it. Indeed, we just need a little more philosophical imagination to come up with a counterexample to this modified account.

Consider a human subspecies that has grown entirely accustomed to living in mountain ranges with steep drops and narrow ridges—a *Homo Monticola*, if you will. Because of their treacherous habitat individual members of this species are scarce and have therefore evolved in such a way that they are unable to intentionally kill one another. However, this species has not lost the ability to form the intention to kill because this ancestral remnant still has a biological role to play. The *Homo Monticola*, namely, lives in small nomadic groups and for safety reasons they have developed a culture of using ropes to hold on to each other when they traverse difficult areas. Sometimes however, things go astray and one member that gets stuck or falls down threatens to endanger the entire group. When that happens the *Homo Monticola* holding on to the unfortunate group member forms the intention to kill that one member in order for the rest to survive. Because it is extremely rare for her to have such murderous intentions this is a strange experience for the *Homo Monticola*, in fact the experience so unnerves her that it makes her hands tremble which cause her to accidentally let go of the rope. Hence, we have a case where a movement is caused in a biologically proper way by an intention that represents it, without this movement's being intentional.¹⁴⁴

¹⁴⁴It might be questioned whether a mental state that can only find execution via a mechanism which involves nervousness can still be aptly called an

The mechanism that has evolved to sacrifice fellow community members via intention formation and nervousness is non-accidental in a robust biological sense, i.e., the capacity has been evolutionarily selected because it propagates survival. However, when a *Homo Monticola* causes the death of one of her fellow tribe members it in no way constitutes an intentional killing. Hence, we can conclude that although the proper functioning or evolutionary strategy might be able to single out mechanisms that are non-accidental in biological sense, it is unable to extract the sense of non-accidentality required for intentional action. And therefore the evolutionary strategy is unable to save the causal theory of action.

Ralf Stoecker's dispositional strategy. Ralf Stoecker thinks that we should let go of the idea that mental states are “gears or clutches in a causal chain leading to the behaviour” [Stoecker 2003, p. 308]. Instead we should think of mental states as dispositions to produce behaviour.¹⁴⁵ When we think of intentional attitudes as causal powers with intentional actions as their manifestations, we can avoid the problem of causal deviance. For there is a difference between the occurrence of a causal power in a given causal chain, and the manifestation of the power:

A child might cry because the movie is scary, yet not because

‘intention’. I think this is indeed doubtful, but note that the one who employs the evolutionary strategy has no resources to exclude this state from being an intention. On her account, all there is to an intention is that it is a representational state that causes behaviour in the way it is biologically supposed to cause it. Saying that it is no intention because it can only be executed via nervousness already presupposes that we have an account of causation in the right way in place that excludes nervousness. But the evolutionary strategy was supposed to precisely deliver an account that excludes devious mechanisms.

¹⁴⁵Stoecker is inspired by certain remarks of Davidson, e.g.: “Beliefs, desires and intentions are themselves causal dispositions. A desire for Abel’s death is (no doubt among other things) a disposition to be caused to cause Abel’s death given appropriate beliefs, the opportunity, etc.” [Davidson 1990, p. 22].

it was frightened, but because it is not allowed to watch scary movies. Here one can [...] say that the child cried because the movie was scary. But [...] this is highly misleading, because it suggests that the crying was an *actualization* of the movie's causal power of being scary, which it wasn't. [Stoecker 2003, p. 310]

In the climber case it would also be true, but similarly misleading, to say that the climber's intention caused him to loosen his grip: although that intention occurred in the causal chain, the climber's movement was no actualisation of the intention.

Of course, this solution to the problem of causal deviance only works if there is a way to spell out what distinguishes the mere effects of a causal power from its manifestations.¹⁴⁶ And even if that can successfully be done, we still need to rule out proper manifestation of intentional attitudes that are not intentional actions (like the production of saliva). Stoecker tries to overcome these issues by arguing that our intentional actions are also manifestations of a "rational causal power" to move our limbs as we want them to. But he offers no further analysis of this power. Unfortunately, this addition renders his account totally useless for the causal action theorist. For CTA aims at a reduction of the agential to the non-agential, so no rational powers to move can be left unanalysed.

Rowland Stout's Aristotelean strategy. Rowland Stout [2010] argues that the reason why philosophers are unable to overcome causal deviance, is that they accept a Humean metaphysics of causation. Hume does not offer the causal action theorist the resources to specify what causation in the right way is. For Hume, a causal *process* is nothing more than a chain of causally related *events*. As Stout argues, this leaves the Humean with only one way of specifying how one event causes another: by specifying the intermediate events between the cause and effect. Now it takes

¹⁴⁶I will present my own account of powers and manifestations in Chapter 2.

“relatively little philosophical imagination to conjure up a new deviant causal chain counterexample in response to any such specification” [p.162]. We can always think of some sort of bizarre mechanism that deviantly connects the newly specified intermediaries. Therefore, Stout argues, the causal action theorist should abandon Hume’s metaphysics of causation in favour of Aristotle’s. On the Aristotelean picture, a causal process is the actualisation of a potentiality. We can less mysteriously express the ‘actualisation of a potentiality’ as: ‘the manifestation of a disposition’, or ‘the operation of a mechanism’. The Aristotelean, then, has a different way of specifying how one event causes another:

According to this approach, specifying the way A causes B is to specify what sort of mechanism it is whose operation B belongs to and that takes A as an input, or is set off by A, or is somehow embodied in A. If an effect belongs to the operation of the right sort of mechanism, then it is caused in the right sort of way. In this approach, a causal process is not understood as a causal chain, so deviant causal chains are irrelevant. [Stout 2010, p. 162]

This is way too quick. True, Stout’s Aristotelean metaphysics leaves no room for any causal chains and therefore by definition rules out those of the deviant variety. But the Aristotelean metaphysician has a task very similar to specifying the right causal chains: she has to indicate the right sort of causal mechanism.¹⁴⁷ Stout is aware that a further specification of the causal mechanisms involved in agency is needed. But he seems to think that this can be unproblematically achieved when we consider that “the characteristic quality of the sort of mechanism whose operation constitutes an intentional action is its guidedness or goal-directedness” [ibid, p. 163].

¹⁴⁷Stout’s strategy is very similar to Stoecker’s solution and, as we can see has a very similar problem. The main difference between the two strategies is that Stoecker is a Humean who believes that we should give a reductive account of dispositions.

This takes us back to many of the strategies discussed above, and as we saw, all of these strategies failed at giving a reductive account of guidedness and/or goal-directedness. As an Aristotelean, Stout might not care too much about reduction, and therefore his account of the characteristic quality of action-mechanisms might fare better.¹⁴⁸ But it is clear that going Aristotelean by itself offers no advantage to the proponent of CTA.¹⁴⁹

What do we learn from all these killings, stampeding boars, nervous climbers and clever neuroscientists? Many of the scenarios put forth as counterexamples to CTA seem pretty far fetched. Of course, even mere conceptual possibility can show that a particular specification of CTA is definitionally inadequate. But without a more fundamental understanding of why all of the proposed accounts fail, it is hard to see why the problem of causal deviance would be more than a “somewhat technical problem” [Enç 2003, p. 3]. A technical problem that some may try to solve with new and more complex definitions, but which can be safely ignored by those who want to concentrate on the philosophical import of CTA instead of on its technical details. Only if we can find a deeper flaw that explains the failure of all the proposed solutions to the problem of causal deviance, can we have a more definitive argument against the causal theory of action.

Surprisingly, philosophers who are critical of CTA have said relatively little about why the deviant causal chains are so hard to overcome. Mayr suspects that actions essentially are “abilities to act at will” [Mayr 2011, p. 140] and that it is impossible to explain these abilities and their manifestations in purely event-

¹⁴⁸For a critique of Stout’s attempt at understanding goal-directedness see [Mayr 2011, p. 133-136].

¹⁴⁹As we will see in Chapter 6 I will also defend a (broadly) Aristotelean metaphysics of powers. However, the problem of causal deviance should be solved by a theory of action, and not by an account of causation. More on this in §3.4.

causal terms. But he offers no deep reason for why an event-causal explanation is untenable and merely points to the failures of the existing analyses. Similarly, Sehon suspects that CTA fails because teleological explanations are irreducible and action explanation is necessarily teleological. But as he admits, “the failure of extant analyses does not demonstrate the irreducibility of teleological explanations” [Sehon 1997, p. 212]. Was this entire section then totally in vain? Can the proponent of CTA simply brush aside the problem of causal deviance? I think such a dismissive attitude cannot be justified, as I will attempt to show in the next section.

3.4 A diagnosis

When the founding father of the causal theory of action and originator of the idea that the mental should cause behaviour ‘in the right way’, revisited the explanation of action later in his life, he wrote:

Several clever philosophers have tried to show how to eliminate the deviant causal chains, but I remain convinced that the concepts of event, cause and intention are inadequate to account for intentional action. [Davidson 2004, p. 106]

Surprisingly, this conviction did not lead Davidson to abandon CTA. What justification could he, and the many who share his dismissive attitude, have for upholding the causal theory? Was CTA not precisely the thesis that the notions of “event, cause and intention” are sufficient for the analysis of action?

Perhaps one can hold on to CTA in the face of the failure to come up with a solution to the problem of deviant causal chains, because one thinks that problem should be addressed outside of philosophy:

[P]recisely what is this ‘characteristic’ mode of causation by

which wants and beliefs cause intentional action? [...] A complete explanation of how wants and beliefs lead to intentional action would require extensive neurophysiological information, and I do not think it is fair to demand of a *philosophical* analysis that it provide this information. [Goldman 1970, p. 62]

But how are neuroscientists going to tell normal causal chains from deviant ones? Neurophysiologically, nothing goes wrong when the climber's nervousness causes his hands to tremble and his grip to loosen and it is perfectly normal to get nervous when one has murderous thoughts. It is only when we start doing action theory that we recognise the oddity in the causation of an unintentional movement by an intention which represents that very movement. As Geert Keil remarks: "Our brains could not care less about deviant causal chains" [Keil 2007, p. 76]. Hence Alvin Goldman is mistaken.

But even if it is philosophy that has to solve the issue of causal deviance, is this a task for the philosophy of action? Rowland Stout, for instance, argues that "it would be a real advance if [the philosophy of action] could pass the problem of deviant causal chains to the philosophy of causation" [Stout 2010, p. 163].¹⁵⁰ A further argument for outsourcing the solution to the problem of causal deviance is that it does not exclusively trouble action theory, but also forms a potential problem for causal analyses of perception, meaning, emotion and knowledge.¹⁵¹ If the intractability of the problem of causal deviance shows that the causal approach to intentional action is mistaken, then it "would also

¹⁵⁰In the last section we saw that Stout himself thinks that the switch from a Humean to an Aristotelean metaphysics of causation is crucial for a treatment of the problem.

¹⁵¹It is not sufficient for the perception of an object, for example, that it is that object that causes you to have a representation of that object. If the object somehow brings about a hallucination of the object, then the object is represented and the representation is caused by the object, but the object is not perceived. These cases are known in the philosophy of perception as cases of "veridical hallucination", see [1980].

show that a causal approach to absolutely anything is mistaken” [Stout 2010, p. 163].¹⁵² To ridicule the thought that all causal theories are flawed, Stout gives the example of a causal theory that surely must be fine: the causal theory of cooking. On that theory we say that the heat of the oven cooks the chicken if and only if it causes the chicken to be in a cooked state. This theory, according to Stout, also falls prey to causal deviance, because the heat could trigger some microwave activity which then cooks the chicken. Stout’s analogy, however, is misguided, for in what sense is this causal chain that runs through the microwave deviant? There can be a perfectly normal causal mechanism that connects the oven and the microwave, and we could even imagine that a microwave manufacturer would design its products in such a way that the waves are always triggered by a heat-generating contraption.

When we exclusively look at the causal processes unfolding nothing strange happens. Waywardness only enters the fray when philosophers try to causally account for so-called *intentional* notions. Intentionality is the quality of *being about*, or *being directed at* something. Intentional action is intentional because it is directed at attaining the agent’s ends. Similarly, knowledge is intentional because it is about something, and perception is always a perception *of* something. Only when we try to give reductive accounts of such intentional notions in terms of causation deviance arises, because the connection between an intentional state and the object it is about, or directed at, can be merely accidental. This happens precisely when the causal connection is not in virtue of the aboutness or directedness of the intentional state. As I will show later in this section there are good arguments to

¹⁵²As we saw above (p. 113), Normy Arpaly [2006, p. 71] voices a similar thought when she argues that the problem of causal deviance also besets those who want to explain how the weight and size of a stone cause the breaking of a window.

think that a causal reduction of intentional agency is impossible, and there might very well be similar arguments against the reduction of other intentional notions. But contrary to what Stout thinks, this does not show that we cannot give a causal account of anything at all. We just have to conclude that reductive causal accounts of *intentional notions* are mistaken. If intentionality is not reducible to causality this would not mean that there is something wrong with causality simpliciter. We can conclude that the problem of causal deviance is not to be solved in the philosophy of action but should be of concern to anyone who wants to reduce intentionality. Therefore it remains puzzling that Davidson brushes aside the problem of deviance with such ease.

Torbjörn Tännsjö [2009] thinks that Davidson can motivate the dismissal of the problem of causal deviance differently. There might not be a *general* solution to the problem of causal deviance, Tännsjö argues, but this does not mean that we cannot distinguish between the correct and wayward causal chains in relation to specific action types. Unfortunately, when he attempts to rule out deviance for specific action types, Tännsjö does not go beyond arguing that certain types of action require continuous guidance, or that others require to be responsive to the content of intentions. As we have seen in the previous section, these notions are useless to the causal action theorist as long as we cannot reductively understand them. Hence, Tännsjö's proposal misses the mark. Nevertheless Tännsjö is right that Davidson did not see the need for a *general* solution to the problem of causal deviance. But equally, Davidson might not have seen the need for individual solutions for specific actions types either.

What then justifies continued allegiance to CTA? We can find a clue in Davidson's own words:

If [CTA] is correct, then acting with an intention does not require that there be any mysterious act of the will or special attitude

or episode of willing. For the account needs only desires (or other pro attitudes), beliefs, and the actions themselves. There is indeed the relation between these, causal or otherwise, to be analysed, but it is not an embarrassing entity that has to be added to the world's furniture. We would not, it is true, have shown how to define the concept of acting with an intention: the reduction is not definitional but ontological. But the ontological reduction, if it succeeds, is enough to answer many puzzles about the relation between the mind and the body, and to explain the possibility of autonomous action in a world of causality. [Davidson 1978, pp. 87–88]

Davidson seems to think that the problem of causal deviance merely shows that we cannot define the concept 'intentional action' in terms of representation and causation. But the lack of a conceptual reduction in his eyes does not imply the impossibility of an *ontological* reduction. Davidson's philosophy as a whole can be seen as an attempt to reductively understand complex entities in a world that is entirely made up out of simple objects and events—his reductive account of meaning is just as, if not more, notorious as his account of intentional agency.¹⁵³ Hence Davidson's motivation for CTA might go well beyond the philosophy of action and rest on his conviction that deep down only objects and events exist. If there are to be such things as intentional actions, we have to somehow understand them in terms of the fundamental entities. This is not the place to argue against such a *Weltanschauung*.¹⁵⁴ But there are good reasons to question Davidson's confidence in the possibility of an ontological reduction of agency.

Recall that Davidson rejected neo-Wittgensteinian accounts of action, because they leave open the possibility that an agent's bodily movement only accidentally conforms to her reasons. Davidson adds a causal connection to make sure that the agent not only acts

¹⁵³See, e.g., Davidson [1967].

¹⁵⁴In Chapter 6 I will offer some arguments against this world view on the grounds that it cannot account for dispositional properties.

in accordance with her reasons but also acts *for* the reasons she has. What the problem of causal deviance shows is not merely that intentional action cannot be defined in terms of causality and representation, it shows that the causal connections Davidson introduces are not sufficient to rule out accidentality. Even when a reason causes the movement it represents, it might still be that the movement was not performed *for* that reason. This is what philosophers who argue that reasons should cause actions in virtue of their content, also pick up on. But why would we think that it is possible to reductively understand such causation in virtue of content?

As David Horst [2012] has recently pointed out, the causal action theorist has to think of representation and causation as logically independent.¹⁵⁵ It is a feature of CTA that there is no essential difference between the kind of causality exhibited in intentional action, and the kind of causality found outside of the realm of human action. It is this feature that makes CTA an explanation of how agency unproblematically fits into the ontology of scientific naturalism. Hence whether a mental state causes another mental state, my mouth to water, or the raising of my arm, the kind of causal connection is always the same. Therefore causation is blind to content—as far as its causal role is concerned my mental state operates independently of its content.

On the picture CTA sketches, representation is equally blind to causation. A mental state can represent something without bringing it about (or being brought about by it). My hope that the San Diego Chargers will win the Superbowl does represent them winning, but unfortunately does not contribute to the actual occurrence of that state of affairs. The independence of representation from causation is especially clear when we consider

¹⁵⁵Something that is nicely reflected in Davidson's remark we came across in section 3.1 that action and mental event are "related in two very different ways" [1982, p. 141].

that proponents of CTA standardly think of mental states as propositional attitudes. On this picture a mental state consists of two independent elements: a proposition (its content) and an attitude towards that content. Hence, my belief that I will have Kung Pao chicken tonight and my desire for having Kung Pao chicken tonight represent the exact same proposition. But both play an entirely different causal role. My desire to have Kung Pao chicken causes me to go and cook it, whereas my belief that I will have the dish is caused by, say my belief that I will cook it.¹⁵⁶ Hence my mental state can represent an object independently of the possible causal relation it has with that object.

Because CTA considers causation and representation to be logically independent it lacks the resources to account for causation in virtue of content. Therefore CTA can never exclude the possibility of accidental causal connections between a mental state and the movement it represents. Horst concludes from this that the problem of deviant causal chains is inescapable for the causal action theorist. I would like to go one step further. The problem for CTA is not just that it cannot exclude deviant causal chains, it is that it cannot explain when a causal chain is not deviant, i.e., when we have a genuine intentional action. If representation and causation are independent, then their co-occurrence can only be accidental. On CTA every causal chain from intention to action is in a sense deviant, because for any intentional action it is a mere accident that it is caused and represented by the same state. Perhaps this is what Anscombe had in mind when she wrote that:

[Davidson] speaks of the possibility of ‘wrong’ or ‘freak’ causal connexions. I say that any recognizable causal connexions would be ‘wrong’, and that he can do no more than postulate a ‘right’

¹⁵⁶Most philosophers of mind (e.g. Searle [1983] and Smith [1987]) even argue that the difference between different kinds of mental state consists in the different causal force they have. In other words, a belief is a belief because it tends to be caused by its object and a desire is a desire because it tends to cause its object. More on this in §5.5.

causal connexion in the happy security that none such can be found. [Anscombe 1974, pp. 378]

We have to conclude that Davidson's attempt to rule out accidentality with the help of causation has miserably failed. The problem of causal deviance merely brings out this deficiency of CTA: it is because representation and causation are independent that we can construe the deviant scenarios. Rather than being a problem that we can easily dismiss as merely definitional, causal deviance is a *symptom* of a much deeper problem. The deep problem is that on CTA it can only be merely accidental when a mental state represents the same movement that it causes. This accidentality goes hand in hand with CTA's assumption that content and representation are independent. But this assumption cannot simply be dropped, since the causal action theorist needs precisely this independence to achieve a decomposition.¹⁵⁷ We have seen that intentionality cannot be reduced to the co-obtaining of causation and representation, but that it requires causation in virtue of representation. Hence the causal theory of action fails to provide an account of agency.

3.5 Concluding remarks

In this chapter we have discussed the two main problems for the causal theory of action. Upon inspection the problem of the disappearing agent turned out to be a mere complaint that CTA is reductive, instead of an argument against such a reduction. The problem of deviant causal chains is much more serious. We have looked at many different proposals to rule out causal deviance and have argued that none of them are successful. This is not

¹⁵⁷For Davidson's own account it is particularly clear that the content and causal role of a mental state have to be independent because on his view there are no bridge laws between the psychological and the physical. See Davidson [1970].

surprising in the light of the last section where we found that the problem of causal deviance is only a symptom of a deeper problem for CTA. CTA cannot rule out accidentality and therefore fails to give a (reductive) account of agency.

What does all of this tell us about free will? In section 3.2 we saw that Pereboom attacked event-causal libertarianism on the ground that it does not properly explain how an action can be *up to* the agent. He seemed to think that the main cause for this lack of control is that event-causal libertarianism is indeterministic. But we can now see that the reason why event-causal libertarianism fails to assign a proper role to the agent in the production of action, is that it unsuccessfully tries to reduce intentional agency to event-causation. CTA simply lacks the resources to explain how an action can be more than a mere accident. Therefore, there is a problem about luck for *any* event-causal story—including the deterministic ones. This is a remarkable conclusion given the fact that luck is usually considered a problem for indeterministic theories of agency only.

The result that CTA does not provide an adequate account of agential control, is especially troubling to the compatibilist. In the last chapter we saw that CTA was inherently compatibilist and that many, if not all compatibilists accept the account of control CTA provides. Hence the result in this chapter—that CTA does not provide the correct account of agential control—forms a strong argument against the compatibilist. Some might even argue that it destroys compatibilism because compatibilism necessarily involves CTA.¹⁵⁸ I, however, do not want to rule out the possibility of different forms of compatibilism in advance. Nevertheless the fact that CTA is such a widespread theory can no longer be used as an argument in favour of compatibilism.

Recall that the aim of this thesis is to understand and defend

¹⁵⁸A large part of John Bishop's [1989] book is concerned with providing an argument for the claim that all compatibilists have to accept CTA.

the simple libertarian picture of free will. In Chapter 2 I argued that the first main task for anyone who wants to understand this picture consists in undermining CTA.¹⁵⁹ That task is now complete. In the next chapter we will move on to the second task and start the search for an alternative account of intentional agency.

¹⁵⁹See p. 85.

Chapter 4

Accidentality in Action II

Non-Causal and Agent-Causal Theories

Willing, if it is not to be a sort of wishing, must be the action itself. It cannot be allowed to stop anywhere short of the action.

—Ludwig Wittgenstein¹⁶⁰

The event-causal theory of action (CTA) has reigned supreme since Davidson first introduced it. Nevertheless in the last few decades there have been some attempts to dethrone CTA. The proposed alternative accounts of agency come in two variants. Some philosophers believe that actions are to be explained without reference to their causal history and therefore propose non-causal theories of action. Others contend that there is nothing wrong with a causal account per se, but that it is wrong to think of mental events as the causes of actions. Instead they believe that agents themselves

¹⁶⁰Wittgenstein [1953, §615].

directly cause their (free) actions, and hence defend agent-causal theories. This chapter will investigate whether these non-causal or agent-causal theories succeed where CTA failed. Are they able to exclude accidentality? Can they give us an account of acting for a reason and thereby explain the kind of agential control required for free will? In section 4.1 I will argue that the main non-causal theories fail to properly account for intentional agency, and that there is good reason to suspect that any non-causal account of agency is doomed to fail. In section 4.2 contemporary agent-causal theories will be considered. They too fail to provide a satisfactory account of agential control. But, I will argue, their failure is not due to a fundamental flaw in agent-causalism. Therefore there is some hope that we can eventually arrive at a successful agent-causal theory of intentional action that accounts for the simple picture of free will.

4.1 Non-causal action theory

The most obvious alternative to a causal theory of action is, of course, a *non-causal* theory of action. But what can such an account look like? In this section we consider four proposals: Julia Tanney's account of reasons as context placing explanations (§4.1.1), Scott Sehon's teleological realism (§4.1.2), Hugh McCann's idea of intrinsically intentional decisions (§4.1.3), and Carl Ginet's account in terms of concurrent intentions (§4.1.4).

4.1.1 Reasons as context placing explanations

When Davidson stated that reasons must be causes, he was explicitly arguing against the neo-Wittgensteinian thought that rationalisation and causal explanation are radically different.¹⁶¹ Therefore it might be helpful for the non-causal action theorist to revisit

¹⁶¹Cf. §3.1.

neo-Wittgensteinian thoughts. This is exactly the route Julia Tanney [2005, 2009] takes. She goes back in particular to the account of A.I. Melden [1967] who is notorious for claiming that causal explanations are “wholly irrelevant” [p. 184] to the understanding of human action—a statement that derives its fame from being dismissively quoted in Davidson’s ‘Actions, Reasons and Causes’. According to Melden, a causally antecedent physical event—like a happening in the brain—always lacks the logical force to turn a bodily movement into an intentional action. CTA, of course, claimed that prior events *can* have this logical force when we understand them as motives, reasons, intentions or believe-desire pairs, i.e., when we understand them as *mental* events. But this, Tanney [2009, p. 96] observes, would not satisfy Melden. Mental events can only help to explain the rationality of the action if they are motives, reasons or intentions *for* the action. However, on Melden’s view, this is incompatible with them also explaining what turns a mere movement into an intentional action:

A motive is a motive for some action either performed or considered; hence a motive, far from being a factor which when conjoined with any bodily movement thereby constitutes an action, actually presupposes the very concept of an action itself.
[Melden 1967, p. 81]

I am not sure whether this argument is fully convincing.¹⁶² But the thought that the relation between an action and the reason *for* which it was done cannot be reduced to causation and representation *by* the reason is familiar enough. As we saw in the last chapter, the failure of such a reduction is the reason why CTA cannot rule out accidentality and thus has to be rejected.

¹⁶²It is in any case hard to see how it would convince the contemporary philosopher of mind who “learned in her first, introductory course on the subject that intentional states enter into both logical relations with other states (in virtue of their content) and causal relations with other states (in virtue of their form)” [Tanney 2009, p. 96].

How then are we to understand agency? And what role do intentions play in the explanation of action? According to Tanney, an agent's mental events often take no part at all in the explanation of action. To make sense of an action, it is sufficient to point to the context in which the action took place. "Why did the woman run out of the building?" "Because it was on fire." The building's being on fire provides a reason for a person's running out of it. Hence our puzzlement as to why the woman fled is fully alleviated by considering the circumstances. This, Tanney argues, is the standard form of action explanation. What the agent believed or desired before or while she acted, is a question that normally does not arise.

But sometimes, Tanney admits, we do mention an agent's motivations or intentions in the explanation of her actions. "Why did the bus driver raise her arm?" "Because she intended to greet her colleague." However, Tanney argues, even when an intention is ascribed, we are not pointing to something inner, hidden from anyone but the agent. Instead ascribing an intention is just a fancy way to put the agent's behaviour in a context that renders it intelligible. By attributing the intention to greet her colleague to the bus driver, we point out that she spotted her colleague in the other lane and we suggest something about the social practice of exchanging niceties among bus drivers. To suppose that there is a mental event designated by the term 'intention' is not just unnecessary, according to Tanney, it also obscures the way in which action explanation functions:

The existence of [mental] occurrences is not required for the concepts of intention, motive, reason and so on to discharge their explanatory role, thus throwing into question the whole idea that this explanatory role is causal. [Tanney 2009, p.79]

Tanney is surely right that our questions about the behaviour of a particular agent can often be answered by considering the

circumstances in which she acted. But is it always sufficient to merely take stock of the circumstances? What if the woman only fled because she thought that the building was on fire while it actually was not? Here it seems that we *do* need to enquire into the contents of the agent's mind if we want to explain her behaviour. Tanney disagrees: ascribing the thought that the building is on fire to the agent, again just provides another re-description of the circumstances. The only role played by the thought-attribution is that it makes salient that the woman smelled her colleague's half-burned lunch or heard a false alarm. Here one might object, however, that there could be cases where we can find nothing in the circumstances to explain why an agent acted as she did. Tanney reluctantly concedes this possibility. In such cases, she admits, we can do nothing but resort to asking the agent what her goal was in acting. But the answer to that question, Tanney stresses, is again constrained by the circumstances. We would not, for instance, accept the woman's answer that she ran from the building because it was on fire, if it turns out that she was a fire fighter on the job. In other words, the agent's answer only provides an action explanation if it renders the action intelligible given what we already know about the circumstances, and hence the answer cannot just be an introspective report about the mental states the agent had before or during the action. But what if there are no contextual circumstances that allow us to check the agent's answer? Even then, Tanney argues, the agent is the only one who can say why she acted as she did, not because "she has access to something—perhaps conceived as an interior movement—that is forever hidden from anyone else, but rather because no one else has any grounds for correcting her" [Tanney 2005, p. 348].

What Tanney's neo-Wittgensteinian account brings out is that we are often perfectly able to explain actions without mentioning mental events. Even when we do mention these events our aim is

not to give causes of the action, it is instead to bring out the rationality of the action.¹⁶³ But do these considerations bring us any closer to understanding agential control? I do not think so.¹⁶⁴ Although Tanney helps us to better understand our practice of action explanation, her account cannot bring us closer to grasping what it is to act for a reason. As we will see, Tanney's account is just as inept at ruling out accidentality as CTA. The theory of reasons as non-causal context placing explanation suffers from an issue that mirrors the problem of deviant causal chains and might perhaps be called the problem of deviant contextual circumstances. Consider Davidson's climber, whom we encountered in the last chapter. He wanted to rid himself of the weight and danger of holding another man on a rope. When he realised that he had this desire, this so unnerved him that he accidentally loosened his hold. Although the climber's behaviour in this scenario is a mere fluke, we are perfectly able to rationalise his movements with a context placing explanation. The climber after all was in circumstances where it made sense for him to loosen his grip. Indeed his situation may have been so dire that letting go of his fellow mountaineer was the only way in which our climber could have survived. Hence we can use contextual circumstances to rationalise movements, like the accidental loosening of the rope, that do not constitute intentional actions.

But even in the case of genuinely intentional behaviour, context-placing explanations can lead us astray. This happens, for instance, when the circumstances in which an agent acts allow us to rationalise her actions in several different ways. The woman might

¹⁶³Stronger, if an agent responds to the question why she moved in a certain way by pointing to an event-cause, this often constitutes a denial of the movement's intentionality. "Why did you knock over that coffee cup?", "Because I was startled by that ring of the doorbell". More on this in §5.1.

¹⁶⁴Tanney herself is mainly concerned with properly describing the practice of action explanation and hence her theory is not aimed at answering this question. Nevertheless it is interesting to see how much we can say about agential control based on Tanney's neo-Wittgensteinian account.

have run from the building because it was on fire, or because she left her car keys in her car. Given the circumstances, we might be equally keen to accept either of those explanations. So how can we decide which reason was *the* reason on which she acted? It gets even worse when one possible rationalisation is more acceptable given the circumstances than a second rationalisation, but where the second rationalisation is nevertheless correct. What if the woman really ran out because of the sudden desire to go and grab a hamburger?¹⁶⁵ Tanney's account does not provide us with any means to identify which possible rationalisations merely justify the agent's action, and which rationalisation gives the actual reason for which she acted. Therefore an agent can be in circumstances that only accidentally provide a rationalisation of her behaviour, e.g., when she runs out of a burning building for some other reason than its being on fire. We must conclude that Tanney's contemporary revival of neo-Wittgensteinianism falls prey to the same criticism that Davidson levelled against its forefather. An account of agency is incomplete without a further link between one of the justificatory reasons for the action and the action itself, "for a person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it" [Davidson 1963, p. 28]. In other words, merely having a justificatory reason when one acts, is not sufficient for the truth of a reason-explanation that cites that reason.

To be fair, Tanney confesses that she is "deeply pessimistic" [2005, p. 340] about the possibility of constructing an account of acting for reasons. But she does not see the need for it either. She argues that Davidson's objection only shows that not every context-placing rationalisation is successful. But it is no part of her account that every re-description of the circumstances should succeed. She merely wants to point out that when an action ex-

¹⁶⁵Of course attributing this desire is here only intended as a re-description of the agent's circumstances.

planation does succeed, it does not require the existence of mental events alleged to constitute the reasons for acting. I think that this dismissal is too quick. Surely there must be an explanation of why certain rationalisations are correct and others are false. As we have seen, Tanney thinks that we must judge which rationalisations are correct by further specifying the circumstances. When this fails, we have to ask the agent for her reasons. However, we only need to accept what she says because no one else has any grounds for correcting her. But why is it that the agent is able, in the first place, to tell us why she acted as she did? It seems she is not just making up a story, hence if she is able to answer she must actually have some aim in mind.¹⁶⁶ Now Tanney, of course, is sceptical that having an aim in mind, is having such and such a mental state. And she is even more sceptical that acting for a reason consists in being event-caused to move by such mental states. But the failure of Davidson's account of reason-explanation does not excuse Tanney herself from explaining what is needed for a reason to be *the* reason for which the agent acted, instead of just a reason that justifies the action. Denying that it is possible to explain this, comes down to denying that we can ever understand how a reason explanation can be true. Scepticism about the possibility of giving an account of acting for reasons thus implies scepticism about acting for a reason itself, i.e., about intentional action. The upshot is that anyone who wants to give a non-causal account of intentional agency has to meet Davidson's challenge: she has to explain in non-causal terms what it is to act for a reason.

4.1.2 Teleological realism

Scott Sehon [2005] thinks he can meet Davidson's challenge. He agrees with Tanney that we are not pointing to inner causes when

¹⁶⁶As we will see in the next chapter, the agent's awareness of what she is doing is the essential feature of intentional action.

we ascribe intentional attitudes. Instead these ascriptions are part of what he calls “common-sense psychology”. Common-sense psychology in his view is neither in competition with, nor reducible to, physical science. Where science concerns itself with causal explanation, common-sense psychology offers an independent but equally legitimate mode of understanding: teleological explanation. The canonical form of teleological explanation is: the agent φ 'd in order to ψ . When an action explanation mentions the agent's motives or desires, this is only done because they help us understand the aim the agent has. Hence, ‘Dominique went to the café because she wanted to meet her friend’ can be easily transformed to the canonical form: ‘Dominique went to the café, in order to meet her friend’. Action explanation thus works by specifying “a future state of affairs towards which the behavior was directed, rather than an antecedent state that caused the behavior” [Sehon 2005, p. 13]. However, the state of affairs at which an action was directed can satisfy more than one desire. Dominique might not only desire to see her friend, but also desire to get out of her office. In itself this is unproblematic; a trip to the café can be in service to more than one goal. But there might also be cases, and this is where Davidson's challenge comes in, where the agent's behaviour can be justified with regard to multiple possible goals of the agent, but in which she nonetheless acts *for* only one of them. Sehon thinks that we can pick out the reason for acting if we attend to certain counterfactual truths about the agent. Suppose that Dominique would have been able to resist the temptation to leave her office if she had not known that her friend was in the café, or that she would have stayed in the office if she had known that her friend would also be there soon. On the basis of these counterfactuals we can conclude that Dominique acted for the reason of meeting her friend, rather than in order to get out of the office. Once we recognise that teleological explanation supports counterfactuals,

we can see how it is that an agent acted for one, rather than the other reason.

But on what grounds do these counterfactuals hold? As mentioned, Sehon thinks that teleological explanations are not reducible to scientific causal explanations. Therefore the conclusions of common-sense psychology cannot be grounded in physical facts about the world.¹⁶⁷ Nevertheless, common-sense psychology is highly successful in allowing us to predict and understand the behaviour of rational agents. And the best explanation for this success is that teleological explanations are literally true. Therefore Sehon thinks that we must accept teleological claims about the goals of agents as “brute irreducible facts” [Sehon 2005, p. 219] that defy any further explanation. Accepting the existence of such facts is at the core of the position he calls ‘Teleological Realism’. Hence, the counterfactuals supported by teleological explanations are grounded in unanalysable teleological facts.

All this is rather unsatisfactory. We were after an analysis of what it is to act for a reason, and now we hear that it is just a matter of unanalysable fact.

Sehon is aware that his proposal lacks explanatory force, but thinks that he can amend this by giving an “epistemology of teleological explanation” [ibid, p. 135], i.e., an account of how we can determine the truth or falsity of proposed teleological explanations. When we theorise about an agent’s behaviour, he argues, our aim is to produce an account according to which the agent is as rational as possible.¹⁶⁸ Hence when we have two conflicting theories about the reasons for the behaviour of an agent “it is unreasonable to believe the one according to which the agent is significantly less rational” [ibid, p. 139]. Sehon develops this

¹⁶⁷Because of this disconnection of the teleological from the physical, Sehon thinks that the issue of (in)determinism does not influence the possibility of free action. He thus arrives at a compatibilist position on free will. See [Sehon 2012].

¹⁶⁸In this he follows Davidson [1970].

thought by explaining how we should analyse the rationality of the agent. But we do not need to go into the details of his account to see where it goes wrong. Agents simply do not always do what is most rational to do. I am currently typing on my keyboard. The colleague across from me might explain this by hypothesising that I am typing in order to write this chapter. And she is correct. But had she made the same assessment a few moments ago when I was also typing, she would have been wrong. Then I was typing in order to send an e-mail with a funny comic to a friend. However, given the upcoming deadline it certainly is most rational for me to be working on my dissertation. We must conclude that the teleological explanation according to which the agent is most rational is not always the one that gives the actual reason for which the agent acts. Thus Sehon's account of how we determine the truth or falsity of teleological explanation is flawed, and hence cannot be used to supply his account with the much needed explanatory force.

The non-causalist, it seems, can only escape the predicament in which Sehon finds himself by denying that teleological explanation rests on metaphysically brute and unanalysable facts. Instead she should give an analysis of what it is to act for a reason in non-causal terms. This, indeed, is what the two traditional non-causal theories of intentional action by Carl Ginet and Hugh McCann try to provide.¹⁶⁹ Both Ginet and McCann are a lot less hesitant about invoking the private contents of an agent's mind in explaining her actions. But as I will try to show below, neither of the accounts succeeds in ruling out accidentality.

¹⁶⁹Ginet and McCann are very clear that their non-causalism is motivated by considerations about free will. Both believe that determinism poses a threat to freedom and think they can escape that danger by claiming that actions are un-caused. Their non-causal theories are attempts to make sense of such un-caused actions. By contrast, Sehon's teleological realism is not motivated by compatibilism. Compatibilism rather is the result of accepting teleological realism. Tanney is not at all concerned with questions about free will.

4.1.3 Intrinsically intentional decisions

According to Hugh McCann [1998], agents act for a reason when they have *decided* to act for that reason.¹⁷⁰ On McCann's view decision making is a type of information processing. It transforms the information presented in our reasons, "so that the scenario of action they embody comes to be intended, for just those features the reasons give as justifying it" [ibid, p. 165]. Hence every intention resultant from a decision is *intrinsically intentional*, i.e., it is never just an intention to ϕ but always an intention to ϕ for reason R . Dominique never intends to just go to the café, she always intends to do so in order to see her friend or in order to leave her office.¹⁷¹ To act for a reason, McCann concludes, is to act on an intention that represents that reason. But recall that McCann aims to give a non-causal account of action, which means that the decision or resultant intention cannot play a causal role in the production of the action. This makes it really easy to come up with examples of deviance. Consider our climber again. This time he is a little bit more malicious: he does not just desire to let his colleague go, but actually decides to do so for the reason that it lessens the weight he has to carry up the mountain. However, just before he can act on his decision, something else, say a falling rock, causes him to accidentally loosen his grip.¹⁷² In this scenario the bodily movement the climber intended to perform happened, and the decision to perform that movement was in place, so on McCann's account it seems that the climber acted for a reason. But we know that the loosening of the grip only occurred accidentally, and thus was no intentional action at all.

¹⁷⁰A precursor/predecessor of this view is defended by Stewart Goetz [1988], although he does not further explain how deciding to act for a reason works.

¹⁷¹Or in order to do both.

¹⁷²Note that because McCann's account is non-causal, we do not even need the decision to cause the accidental loosening of the grip to construct a deviant scenario.

4.1.4 Concurrent intentions

Carl Ginet [1990, 1997, 2008] agrees that a movement can never derive its intentionality from an intrinsically intentional decision that happened before the movement.¹⁷³ Instead, he believes that an agent acts for a reason if her movements are not just preceded by a desire or intention to attain a goal, but also are accompanied by a concurrent second-order intention. She has to intend of her movements that they attain the aim specified in the mental state that preceded her action. Here is an example: suppose Sarah wants her glasses that she believes are in the bedroom. If Sarah does enter the bedroom we can say that she did it to get her glasses only if she concurrently intended of her bedroom-entering that by it she would get her glasses. The content of the concurrent intention specifies the reason for which the action was performed. Hence, Ginet can distinguish between reasons that merely rationalise an action, and the reason(s) for which the action was performed. The latter, in contrast to the former, figure in the second-order intention possessed by the agent during her act.

Ginet's theory of acting for reasons is, I believe, hopelessly complex. Do I really have to intend of my finger movements that they satisfy my desire to write this chapter? I am most certainly not aware of such an intention—I just move my fingers, it seems, because I intend to write this chapter. Also, we never seem to cite these second-order intentions in action explanations.¹⁷⁴ But the real problem with Ginet's account is that it also allows for scenarios of deviance. Randolph Clarke [2003, p. 22] suggests such

¹⁷³See his review of McCann [1998]: Ginet [2000].

¹⁷⁴Another problem is that Ginet's account leads to a potential regress. Intention formation itself can be explained by reasons. 'Why did Sarah intend to get her glasses?', 'To go and watch the football game'. As Clarke points out: "Since Ginet's account of the reason-explanation of an action appeals to [a further second-order intention], the question arises what can be said about the reason-explanation of the acquisition of that intention. Repeating the same sort of account here would generate a regress" [Clarke and Capes 2014].

a scenario. Suppose that Sarah not only wants her glasses, she also has a desire for her friend Ralph's company. Ralph just came home after a hard day of work and is taking a nap in the bedroom. Sarah knows this and therefore she also wants to let him sleep. Now suppose Sarah enters the room with the concurrent intention that by entering she satisfies her desire to get her glasses. However as it turns out, her desire to get her glasses plays no role at all in causing her entry, while her desire to wake Ralph, of which she is fully aware when she acts, does play such a role. Is the content of her concurrent intention authoritative about the reason for which Sarah acts? Or is she fooling herself? In his response to Clarke's challenge Ginet basically capitulates:

I have to admit that, since it is stipulated that Sarah's desire to wake Ralph plays a role in causing her act of entering, it will certainly seem right to say that she entered *because* she had a desire to wake Ralph and even that her desire to wake Ralph was one of her reasons for entering. [Ginet 2008, p. 232]

According to Ginet this shows that the having of a concurrent intention is not (contrary to what he previously defended) necessary for establishing the truth of reason-explanations. He still holds, however, that there are cases where it is *sufficient*. Therefore, he argues, contra causal action theory, that we do not always have to cash out rationalisation in causal terms [ibid, p. 236]. Ginet is right about this, just like Tanney was right that rationalisations often point to circumstances instead of causes. But by retreating to such claims Ginet and Tanney both just reiterate what is wrong with CTA, where we are searching for an alternative theory of acting for a reason.

No doubt, clever non-causal theorists can try to amend their views with extra conditions to avoid the scenarios of deviance encountered above. There are, however, good reasons to think that a

non-causal account can never succeed to fully rule out accidentality. Just like the failure to rule out deviant causal chains is a mere symptom of a deeper problem for CTA, the scenarios of accidentality for non-causal theories point to a deeper issue with them. CTA fails to rule out accidentality because it tries to *separately* account for the rationality and occurrence of an action—by means of representation and event-causation respectively.¹⁷⁵ Non-causal theories fail because they have no story about the occurrence of actions whatsoever. If Davidson was right about one thing, it is that there is more to actions than their rational structure. Actions do not only have a place in a teleological structure, they also are material happenings. The challenge for any theory of action is to account for this dual aspect. How can one thing be both rational and material? By only focusing on concurrent intentions, decisions, brute teleological facts or rationalising circumstances, we can never hope to understand how actions actually happen in the world. Therefore we should be suspicious of any non-causal account of intentional action.

4.2 Agent-causal accounts

In the previous section we have seen that non-causal theories of action do not constitute a proper alternative to the event-causal theory of action. This section will examine whether agent-causalists can do better. Most agent-causal accounts are not directly focused on the explanation of intentional action. They rather want to explain free action.¹⁷⁶ Unlike non-causal accounts of agency, agent-causalists do attempt to account for the material happening of free actions. On their view, free actions, or the volitions lead-

¹⁷⁵See §3.4.

¹⁷⁶Of course, I argued in Chapter 2 that it is a mistake to think that an action can be intentional but not free. As we will see later in this section, this mistake is a basis for the failure of some of the contemporary agent-causal theories.

ing up to them, are directly brought about by the agent. Agent-causation, they believe, cannot be analysed in terms of causation by agent-involving events such as her beliefs and desires. Such non-reductive agent-causation is especially appealing to libertarians about free will because it can nicely capture and explain the thought that not everything is predetermined.¹⁷⁷ An agent, it is argued, is not the kind of thing that can itself be caused (although events involving the agent, like her birth, can be). Therefore when an agent causes her action or volition, she is “in a strict and literal sense” [Clarke 2003, p. 134] the ultimate source or uncaused cause of the event. Agents can form the fresh starts of new causal chains.¹⁷⁸

Of course the major challenge for the agent-causalist is to explain her irreducible notion of agent-causation. Many philosophers think that the concept of agent-causation is utterly impenetrable. Appealing to that concept to clear up the mystery of free will seems nothing more than “giving a name to a mystery” [Nagel 1986, p. 115]. Agent-causation is even said to be “more puzzling than the problem it is supposed to be a solution to” [van Inwagen 1986, p. 151]. And Strawson presumably had the agent-causal story in mind when he famously scolded libertarians for having to resort to “obscure and panicky metaphysics” [Strawson 1962, p. 27]. Adding to the obscurity is the fact that agent-causalists traditionally have alluded to the nature of agent-causation by contrasting it with the causality exhibited in the inanimate world.¹⁷⁹ More

¹⁷⁷Recently, however, it has been argued that agent-causation also is the solution to many of the problems faced by compatibilists. See Markosian [1999, 2012] and Nelkin [2011].

¹⁷⁸See Steward [2008].

¹⁷⁹Thomas Reid, who developed an early modern theory of agent-causation, for example writes in a letter to Lord Kames: “I am not able to form any distinct conception of active power but such as I find in myself [...] But, if there is anything in an unthinking inanimate being that can be called active power, I know not what it is, and cannot reason about it” [Reid 1895/1983, p. 59]. Richard Taylor offers a similar sentiment: “causation by agents—is so different from the kind of causal sequence found in events that it is unfortunate,

recently, however, the advocates of agent-causation have claimed that the phenomenon is much more ubiquitous than it might *prima facie* seem. Let me introduce the main contenders.

Tim O'Connor [1995a, 2000, 2009a] defends a view of agent-causation in terms of powers. An agent exercises her power to act, just like water can exercise its power to dissolve salt. The major difference between the water and the agent is that the water needs a stimulus (an event-cause) to exercise its power (it needs to come in contact with salt), whereas the agent can spontaneously exercise her power. This spontaneity of course might still be thought to be obscure, but the point we should currently note is that the agent-causation exhibited in free acting is no more mysterious than the causation involved in inanimate processes like the dissolving of salt.

Randolph Clarke [1993, 2003] rejects the powers account of agent-causation. Rather, he adopts a view on which causal laws are grounded in relations among universals.¹⁸⁰ On this view a particular event is the cause of another if there is a general causal law that holds between the event-types of which the particular events are instances. We do not need to go into the details of this account. We merely need to note that Clarke argues that causal laws cannot just be instantiated by events, but can also hold between agents and their actions. Therefore there is nothing mysterious about agent-causation: the only differences between agent-causation and event-causation are the causal relata and the form of the laws that relate them.

Jonathan Lowe [2008], and following him, [Steward 2012a], take things one step further. They believe that agent-causation does not just exist next to event-causation, instead they believe

and the source of much error, that we use the same word for both" [Taylor 1973, p. 262].

¹⁸⁰This view is due to David Armstrong [1983] and Michael Tooley [1977, 1987].

that all causation fundamentally is agent-causation (or better, substance causation). They believe, like O'Connor, that causation consists in the exercise of causal powers. Since only substances (and perhaps fields or energy) can have causal powers, only they can be causes. Event-causation can be reduced to substance-causation.¹⁸¹ When we say that one event causes another, this is to be understood in terms of one power-exercise that brings about another. For instance, when we say that an explosion caused the collapse of the bridge, we mean that the exploding *of the bomb* caused the collapsing *of the bridge*. Similarly, when an agent brings about the movement of her arm, she does so by willing it to move (Lowe), or by moving it (Steward).¹⁸² The agent's exercising her volitional power causes the arm to exercise its power to move. Without a further account it might still be deemed mysterious how an agent can move her body, or how willing can effect changes in the world. For now, however, the point again is that the causation involved in the action of human substances is not more mysterious than that involved in the action of non-personal substances.

Whether the accounts of the contemporary agent-causalists really make agent-causation more acceptable as a natural phenomenon depends on the tenability of the metaphysics they propose. However, this is not the place to assess the metaphysical commitments of these theories.¹⁸³ In this section, I want to focus on whether these contemporary agent-causal accounts can provide an adequate account of agential control. With regard to that question, the idea of agent-causation as a ubiquitous natural phe-

¹⁸¹Ruth Groff is another philosopher who has recently defended that understanding causation in terms of powers is beneficial to the free will debate. See Groff [n.d.].

¹⁸²Steward is adamant that 'movings' by the agent should not be identified with movements of her body. An agent's moving of her body, is the agent's causing her body to move. See [Steward 2012a, pp. 30–35].

¹⁸³More on the metaphysics of powers in Chapter 6.

nomenon introduces a difficulty. If inanimate, impersonal entities can also agent-cause (or, more appropriately, substance-cause) effects, then how can we account for specifically *agential* control? We do not want to say that water controls the dissolving of salt, or that the bomb controls the collapse of the bridge, in the same way as a free agent controls her actions. As I have argued in Chapter 2, to understand agential control we have to understand intentional agency.¹⁸⁴ And indeed most agent-causalists think that it is necessary to supplement their theory with an account of acting for a reason in order to fully understand free will. Agent-causation is supposed to provide their accounts with indeterminism and ultimate origination, and the supplemental account of acting for a reason should explain active agential control.¹⁸⁵ In what follows we will look at what the most prominent agent-causal accounts have to say about acting for a reason in order to assess whether they are capable of ruling out accidentality. Can they finally deliver the luck-free account of agential control we need in order to understand the simple picture of free will? But before I address this question, we have to note that not all agent-causalists think that their story has to be supplemented with an account of acting for reasons. Helen Steward in particular has argued against this tenet. Therefore we will start by considering her account.

4.2.1 Steward's agency incompatibilism

There is much to like about Helen Steward's [2012a] book *A Metaphysics for Freedom*. For one thing, Steward correctly argues that

¹⁸⁴Cf. Ruth Groff [n.d., p. 21], who calls understanding how agents can act for a reason "the most salient" issue for anyone who wants to understand free will on the basis of causal powers.

¹⁸⁵Hence these agent-causalists buy into the problematic picture of free will as intentionality+*X* discussed in §2.3. For them the extra ingredient that turns an action into a free action is that the action is agent-caused. Cf. O'Connor [2011, p. 311]: "Agent causation is a necessary feature of freely chosen activity, even though there may be possible forms of intentional activity that lack it altogether".

freedom is not just the absence of coercion or the ability to act in accordance with one's personal character. Although both coercion and character may play a role in higher (socio-political) conceptions of freedom, there also exists an intelligible conception of freedom on the metaphysically basic level. Agents (including animals) can move their own bodies. The difference between being moved by something else and moving yourself, is that when you move yourself, your movements are up to you. According to Steward, the idea that something can be up to an animal is incompatible with determinism. For to say that something is up to an animal is to say that the animal can settle that matter. And, Steward argues, it is impossible to settle something that has already been settled before.¹⁸⁶ Therefore, the past must leave open things for the agent to settle, if there is to be such a thing as agency at all. Thus, Steward is not just an incompatibilist about free action, she believes that agency itself is incompatible with determinism. This makes her account unique in that it moves away from the standard libertarian view of free will as intentionality+*X*. According to that view it primarily is the extra feature which turns intentional action into free action that is incompatible with determinism. I have already argued against this view in Chapter 2 and suggested that reasonable libertarians should think that the control required for intentional action itself is incompatible with determinism. But Steward's so-called 'Agency Incompatibilism' goes one step further. It is her view that not just intentional actions, but all exercises of agency require indeterminism:

[T]he concept on which I shall centrally rely—the concept of settling—can be more readily disambiguated and seen for what it is in a context in which it is being allowed that there are certain actions that are not intentional under any description at all [...]

¹⁸⁶I will not argue for or against Steward on this point. But in response to Steward's book, many authors have defended compatibilist accounts of settling. See, for instance, Broadie [2013], Clancy [2013] and Pereboom [2015].

Thus, when I absent-mindedly scratch my head, slightly shift position in my chair, or jiggle my foot as I type, I shall count as having acted on the conception of action in which I am interested, for although I am doing none of these things intentionally, it seems right to say that they are all movings of my body by me [...] They therefore count as settlings by me [Steward 2012a, p. 34].

Was I wrong to insist that understanding free will equals understanding *intentional* agency? Can we fully grasp the simple picture of free will by merely considering animal self-motion? One thing about which Steward might well be right, is that all self-movement and not merely intentional self-movement requires indeterminism.¹⁸⁷ But I believe that Steward's account of agency in terms of settling is not developed enough to provide a full-blown metaphysics for freedom, instead of a mere argument for indeterminism. The simple picture of free will does not only require that multiple alternative possibilities are open to us, nor is it sufficient that we cause the coming about of one possibility instead of another, rather we shape the future by consciously selecting one option. Consider that some things we settle by moving our bodies do not seem up to us in the right sense. If, unbeknownst to me, there is a gas leak, my striking of a match might settle the occurrence of an explosion. But was the explosion up to me? In one sense, yes: without me it would not have come about. In a stronger sense of up-to-usness, however, the answer is no, because I had no conscious control over the occurrence of the explosion, it was a mere unforeseen side-effect of my action. Now if everything we settle was only up to us in the sense in which the gas explosion is, we would not be free. If settling the undetermined was all there is to free will, then a radium atom that by decaying settles the clicking of a Geiger counter would also be free.

¹⁸⁷A little more on this in §5.3.

To be fair, Steward, realises that she needs to say more about the nature of settling in order to “sustain the idea that an animal may be truly in charge of what it does” [Steward 2012a, p.198]. This is where agent-causation enters the picture. In elucidating the notion of agent-causation, Steward is mainly concerned with explaining how the entire animal, rather than the micro-events occurring in it, can settle matters. Her view is that agent-causation should be thought of as *top-down* control. An agent’s input does not happen prior to the neuronal processes that initiate her movement, nor is it identifiable with these neuronal processes. Rather, the role of the agent is to monitor and control the entire process of action. This monitoring causal role cannot be reduced to the causal roles of the lower level entities within the agent. In order to render this idea plausible Steward argues that top-down causation is present everywhere in nature:

For example, a cell is a structure that, once formed, can be a source of control over the chemical processes that go on within it in the sense that laws and principles that belong to the level of the cell overtake those that belong to the level of the molecule when it comes to understanding how those lower-level processes are integrated and harmonised to serve the purposes of the cell. [ibid, p. 245]

The problem with this approach is that if being in charge of what you do consists in top-down control, and a biological cell can also exhibit top-down control, then it again becomes unclear how we should account for the special up-to-usness involved in agency. How are we to distinguish agency from the top-down control that is exercised by other entities? Although Steward does not give an elaborate answer to this question, she does suggest that agency is distinctive in that it consists in the exercise of a “two-way power”. When an agent acts, she also always has the power not to act: to refrain from acting. [ibid, pp. 156, 160] This suggestion by itself, however, does not help very much in the absence of a further ac-

count of what it is to refrain from acting. Without such a story, it is unclear why a radium atom would not possess the two-way power to decay or not to decay. Especially since Steward believes that it is “essential to avoid any characterization of refrainment according to which it has itself to be a deliberate act” [Steward 2012a, p. 156], hence we cannot exclude the radium atom on the ground that it cannot deliberately choose not to decay. Therefore, admirable as Steward’s agency incompatibilism might be, it ultimately fails to explain what is distinctive about the control exhibited in agency.

Since Steward cannot account for agential control, her account does not supply the full story needed for a metaphysics of free will.¹⁸⁸ Perhaps this is not surprising if we consider that free will requires that we not just *settle* which possible alternative future becomes actualised. We *choose* how we shape the future. Choice must happen on the basis of reasons, for else our choices would not be up to us, they would be random selections that are a mere matter of luck. We must conclude that we cannot properly understand agential control if we only consider self-movement instead of intentional self-movement: self-movement for a reason. And of the latter Steward offers no account.¹⁸⁹

4.2.2 Clarke’s integrated agent-causal account

Randolph Clarke recognises that any account of the causation of action is incomplete without an understanding of acting for reasons. However, he does not see how reasons could play a role in acting if action were only a matter of agent-causation. As we have seen above, agent-causation is attractive to libertarians because it

¹⁸⁸Although her account might still give a good argument for indeterminism.

¹⁸⁹When she does mention reasons it is merely to stress the negative point that they play no causal role in the production of action: we are not caused to act by reasons, we act on them. But no further account of how we act on reasons is provided, see Steward [2012a, pp. 144–154].

explains how an agent can be the ultimate originator of her directly free actions. But the idea that nothing but the agent herself causes her to move is in tension with the thought that agents act for a reason. What role can reasons play in the production of action if it is not a causal role? One option is to argue that reasons merely play a non-causal rationalising role. But as we have seen in section 4.1, all non-causal accounts of acting for a reason are doomed to fail. Therefore Clarke argues that we must supplement the agent-causal account of the causation of action with an event-causal account of acting for reasons. The result is an “integrated agent-causal account” [Clarke 2003, pp. 133–150].¹⁹⁰

When an agent freely acts she exercises her agent-causal power to bring about her movement. This secures that the agent is the ultimate source of her free action. At the same time the action is also indeterministically caused by the agent’s reasons in the ordinary event-causal way. This, in turn, secures that the produced movement is performed for a reason, and thus intentional. Now it may seem that the action is overdetermined if it is separately caused by both the agent and the agent’s reasons. But Clarke thinks that this is not the case. He argues that the reason and the agent are not individually sufficient for the production of the action, unlike for example two bricks that are simultaneously thrown at a window where each of them would individually ensure that the window breaks. Rather, he likens the production of action to cases of causation where two separate causes are only jointly sufficient for the occurrence of their effect: “an increase in temperature and an influx of oxygen may both be needed, in a certain situation, to cause combustion” [ibid, p. 146]. That both the temperature increase and the oxygen influx are needed for combustion

¹⁹⁰Although Clarke proposes this account, he is in the end unsure about the tenability of agent-causation: “We should doubt the possibility of agent-causation, but we should not be very certain about the matter.” [Clarke 2003, p. 210]

is grounded in natural law. Similarly, Clarke argues, it is a matter of nomological necessity that the exercise of agent-causal power always occurs in parallel to event-causation by reasons.

Does this give a satisfactory answer to worries about overdetermination? I doubt it. Without further explanation it seems quite mysterious why there would be laws that ensure the going together of event- and agent-causation in the case of free action. Furthermore, in contrast to the case of combustion, there seems to be an asymmetrical relation between the reasons and the agent: “it is because the agent had those reasons that he (*qua* agent cause) caused the action that he did, not the other way around” [Jacobs and O’Connor 2013, p. 181]. But this is not the place to discuss the issue any further. For we have already established in Chapter 3 that a proper event-causal account of acting for a reason cannot be had.¹⁹¹ Therefore event-causalism cannot come to the rescue of the agent-causalist. Hence, the question how the proponent of agent-causation can assign a proper role to reasons remains.

4.2.3 O’Connor’s agent-causal powers

Timothy O’Connor’s account of agent-causation is different from Clarke’s in the following respects. Where Clarke holds that the agent directly causes her entire action, O’Connor believes that the agent only causes an action-triggering intentional event. The action itself is subsequently caused by the event-causal chain that runs from intention to movement:

My directly causing events internal to myself is my activity par excellence. So, in the case of an observable bodily movement such as waving my hand, my action consists of the causal relation I bear to the coming-to-be of the state of determinate intention to

¹⁹¹Clarke does mention that he thinks that the problem of causal deviance is solvable, but he does not want to dwell on the issue “to focus on matters more directly concerned with libertarianism” [Clarke 2003, p. 26].

wave my hand, plus the sequence of events that flows from that state. [O'Connor 2000, p. 72]

The action triggering intention thus forms an intermediate step in between the agent and her action. This intermediate step offers O'Connor an easy way to include reasons in his account. On O'Connors view, the intention is not just an intention 'to *A*', but its content is 'to *A* for reason *R*' [O'Connor 2002, p. 351], hence the intention is itself intrinsically rational. However, to reap the benefits of this advantage over Clarke's account, O'Connor has the tough job of explaining (1) how the triggering intention can transfer its rationality to the movements it brings about and (2) how the intention acquires its rationality from the reasons possessed by the agent.

With regards to (1), O'Connor does not have much to offer. He merely notes that the way in which the intention must cause and sustain the action "must be of the right sort, not involving any 'deviant' causal byways through remote-control hidden devices installed and operated by mad scientists, or other such everyday foul-ups to which philosophers of action have given attention" [O'Connor 2000, p. 88]. He continues by noting that the task of specifying the right way is in no way peculiar to the theory of agent-causation. Therefore he believes that he can suffice by pointing to some of the possible solutions we have already discussed in section 3.3. But should the agent-causalist really want to be in the same boat as the event-causalist when it comes to causal deviance? No, since we have strong reasons to think that the problem is insurmountable. That O'Connor's account has this event-causal part already allows us to conclude that he cannot properly rule out accidentality and therefore fails to give the right account of agential control. However, it will still be instructive to consider O'Connor's explanation of (2). For it might still be useful for the agent-causalist if O'Connor can explain how an agent-caused event

can be rational. After all, triggering intentions that event-cause actions do not have to be an essential part of every agent-causal account. O'Connors explanation of (2) might turn out to be applicable to other candidates for directly agent-caused events, such as Clarke's 'directly free actions' or Steward's 'movings'.

In his earlier work, O'Connor tried to explain the link between the agent-caused intention and the prior motivational states of the agent in non-causal terms:

I am in agreement with Ginet that the part of the explanation that involves a connection between the prior desire and the present intention need not be causal in nature (apart from the causal connections involved in continuing to have the desire), but may, rather, be wholly internal (similarity of content) and referential. [O'Connor 1995a, p. 193]

The idea is that an intention is formed for a reason if the agent was aware of that reason just prior to the formation of the intention. If my intention to pick up the phone and order Kung Pao Chicken is preceded by a desire for Kung Pao Chicken, then I intended to pick up the phone for the reason of getting Kung Pao Chicken.¹⁹² It is easy to see why this will not do. If the intention is a mental state then it could in principle also be induced by a mad neuroscientist. The content of the inserted intention might still correspond to certain desires the agent had just prior to the neuroscientist's intervention, but the intention would not be formed for a reason. I might feel a sudden craving for Kung Pao Chicken, but this would never lead me to form an intention to order it because I also realise that I had enough meat this week and should therefore stick to vegetarian food. However if the neuroscientist had intervened right after I felt the craving, then, absurdly, on O'Connor's account I would intend and subsequently act for a reason.

¹⁹²O'Connor, like Ginet, also adds the further condition that you continue to desire Kung Pao chicken during the action and concurrently intend of your action that it satisfies that desire.

In later writings O'Connor rejects the non-causal account [2000, 2009a].¹⁹³ Instead he argues that the agent-causation of immediately executive intentions is probabilistically structured by (among other things) the agent's reasons. On this view, the agent's reasons have a causal influence on the agent's capacity to form the intention to act by altering the propensity of her exercising that capacity. This, O'Connor argues, explains how intention formation can proceed on the basis of reasons.¹⁹⁴ However, an agent's reasons do not directly activate the power to form an intention:

On the view I have described, nothing other than the agent himself activates the causal power in this way. To say that I have an objective probability of 0.8 to cause the intention to join my students at the local pub ensures nothing about what I will in fact do. I can resist this rather strong inclination as well as act upon it. [O'Connor 2009a, p. 213]

O'Connor of course wants to say that only the agent can activate her causal power to secure the main feature of agent-causalism. But at the same time the claim that we can resist or act upon our inclinations seems to sever the connection between reason and intention O'Connor's account was meant to provide. To save agent-causation, O'Connor has to say that nothing but the agent can exercise her agent-causal power, but this independence makes it obscure how reasons *do* influence the agent to exercise his power.¹⁹⁵ As Pereboom [2014, p.61] notes, it seems an "unexplained coincidence" that when an agent has more reason to act, she has a higher probability to, of her own accord, exercise her agent-causal power.

¹⁹³Albeit mainly because reasons differ in motivational strength, which cannot be accounted for in non-causal terms. See O'Connor [2011].

¹⁹⁴A reason might structure the power to act by influencing the probability of its exercise even non-consciously. In such cases, O'Connor holds, the agent acts *on* a reason. The condition of sameness in content between the intention and a prior reason-specifying mental state remains necessary to make the reason one *for* which the agent acts.

¹⁹⁵If O'Connor would claim that an agent's reasons indeterministically cause what she agent-causes, then his account would collapse to event-causalism.

Accidentality again rears its ugly head. However, O'Connor does have something to say about how reasons influence the power to produce intentions. He suggests that reasons themselves are also powers: "reasons are, in part, powers to act on the persistent agent causal power, to alter its strength" [Jacobs and O'Connor 2013, p. 182]. But how is this supposed to work? We can understand powers as entities that bring about events or processes, either when they are triggered (such as water's power of dissolution), or spontaneously (such as radium's power to decay). But how do powers alter other powers? A glass can be made less fragile with the help of thermal tempering processes, so perhaps in that case we can say that the heating power of the furnace altered the glass's fragility. However it is unclear, on O'Connor's account, by means of what sort of process reasons alter the persistent agent-causal power.

An alternative explanation of the role reasons play in the exercise of the agent-causal power is offered by Erasmus Mayr [2011].¹⁹⁶ He agrees with O'Connor that having a reason involves a power, and that this power is separate from the core agent-causal power to which Mayr refers as "the ability to act". However:

It is important to see that [the power involved in having a reason] is not itself a power for physical change [...] Therefore, any manifestation of the specific power involved in acting for a reason must, at the same time, also involve a manifestation of another power of the agent—especially an ability to act. [ibid, p. 295]

Mayr's idea is that when one acts intentionally, this is the mutual manifestation of a power to act, and the power of the reason. Unlike O'Connor's idea of one power acting upon another, the idea of mutual manifestation is understandable and familiar from a powers metaphysics of the inanimate world. When water dissolves

¹⁹⁶Mayr's account of agent-causation mainly differs from O'Connor's in that he believes that agent-causation does not have to proceed via the exercise of a spontaneous power. Instead the agent-causal power can itself be triggered by prior events. Hence Mayr's version of the agent-causal theory does not imply libertarianism.

salt, for instance, we have a mutual manifestation of water's power of dissolution, and salt's power of being dissolved. But in that case both powers are powers for physical change, despite of their inability to manifest independently. Mayr's power that isn't a power for physical change, on the other hand, seems quite mysterious. How can something be powerful if it cannot alter anything real? The positing of such a power seems to reintroduce the obscurity that contemporary agent-causalists tried so hard to eradicate.¹⁹⁷

Hence O'Connor's agent-causalism remains in need of an account of how reasons can influence the agent-causal power without taking that power away from the agent.¹⁹⁸ Because we lack such a theory, O'Connor's account fails to rule out accidentality and hence we have to look further for the right account of agential control.

4.2.4 Lowe's volitionalism

Jonathan Lowe believes that all causation is fundamentally substance causation. Although this might help to make causation by human substances less mysterious, he realises that there also must be something special about the way in which an agent causes her actions. What is special about free will, according to Lowe, is that it consists in the exercise of *the will*. The will is a power to produce volitions or willings. Volitions, according to Lowe, are non-physical events. Nevertheless Lowe holds that agents can make their bodies move by willing.

How is the power of the will related to acting for a reason? Lowe argues that the will is an intrinsically *rational* power. By

¹⁹⁷It might be argued that the agent-causal power is a power for inner *mental* change of some kind and that it thus *does* alter something real. I am however unsure whether this makes agent-causal power any less mysterious.

¹⁹⁸That O'Connor fails to deliver such an account is not surprising given the fact that he thinks that actions can be intentional without being free. Because of this, his account of what makes an action rational is necessarily independent of his story about what makes it free.

this Lowe means that the will is “a power whose exercises are responsive to reasons, or which is exercised “in the light of” reasons” [Lowe 2013, p. 177]. Its rationality is what distinguishes the will from, say, the power of radioactive decay: “the atom does not decay ‘for a reason’, or ‘in the light of’ its beliefs and desires, for it has none” [Lowe 2008, p. 176]. It seems correct to say that rationality is what distinguishes specifically agent-causal powers from the causal powers of impersonal substances. Indeed, I too have been arguing that understanding free will comes down to understanding how an agent can act for a reason. However, it is not enough to merely stipulate that the will is a rational power. We want to understand *how* a power can be exercised for a reason. To this question Lowe claims to have a straightforward answer:

[T]he reason for which the agent acted is the reason which the agent chose to act upon. Normally, the agent himself will be able to tell us which reason this was [...] By choosing to act in a certain way in the light of a given reason, the agent makes that reason the reason for which he acted on that occasion. [ibid, p. 183]

This explanation of what it is to act for a reason, seems eerily similar to an account we came across in section 4.1.3: the non-causal account proposed by McCann. McCann also argued that choices (decisions) are intrinsically rational (intentional) because they are made for reasons. The problem for McCann’s account was that it could not explain how the rationality of the decisions transfers to material movements, since he explicitly holds that decisions cannot play a causal role in action production. At that point Lowe would disagree. He believes that it is very common that the exercise of a non-causal power has a causal effect. When a rolling boulder, for instance, pushes a tree, the boulder causes the tree to move by rolling into it. And the boulder rolled into the tree by manifesting its power to roll. But according to Lowe the power

to roll is not a causal power. “It is not a matter of the boulder’s *causing* anything” [Lowe 2008, p. 127], it rather consists in the boulder changing its position.¹⁹⁹ Just as the boulder can cause the movement of the tree by exercising its non-causal power of rolling, an agent can move her body by exercising her non-causal power of willing. One can doubt whether rolling really is non-causal and whether much sense can be made of the causal effects of exercising non-causal powers in this way.

But suppose Lowe’s metaphysics of causation is coherent, then it still does not satisfactorily explain the production of action.²⁰⁰ Even if the rolling of the boulder consists in the exercise of a non-causal power, rolling still is a *physical* process and rolling into a tree is still a physical event. But volitions, according to Lowe, are *non-physical* events. And how can a non-physical happening cause the occurrence of a physical bodily movement? This is something Lowe admits he cannot explain:

I would not pretend to have explained how a person can, by willing, effect physical changes in his or her body—and maybe this is something that is destined forever to remain a mystery to us [ibid, p. 178]

But perhaps one can disagree with Lowe that volitions are non-physical, and argue that volitions are physical events in the brain. Then, willing would be a physical process just like rolling is. And it might have material effects, just like rolling has. The problem with modifying Lowe’s account in this way, is that even if these material willings cause the exercise of the power to move the agent’s body, nothing guarantees that they do so non-deviantly. The willing might cause a nervousness that in turn causes the agent’s body to

¹⁹⁹For Lowe, a power is only causal, if its exercise consists in the causing of some relevant sort of effect to an object. He offers water’s power to dissolve salt as an example of a causal power. See Lowe [2008, pp. 149–151].

²⁰⁰As remarked above, more in-depth discussion of the metaphysics of powers will follow in Chapter 6.

move.²⁰¹ Hence it seems to me that the prospects for a volitionist account like Lowe's are bleak. It either has to construe volition as a non-causal immaterial power, but then it becomes hard to explain how the will can issue in material bodily movements. Or it has to regard volitions as the material causes of bodily movements, but then it fails to explain why these movements are rational.

All the contemporary agent-causal accounts discussed in this section fall short of accounting for agential control. But is there a common reason for their failure? I think there is. The agent-causal theories in fact make a mistake similar to the mistake of the event-causal theory of action. That theory could not rule out accidentality because it separated the rational from the material by accounting for the former in terms of representation and the latter in terms of causation. Now all the agent-causal theories we discussed also separate their agent-causal story about the material production of bodily movement from their story about the rationality of these bodily movements. This is most clear in Clarke's and Steward's accounts: Clarke explicitly denies that we can understand acting for a reason in agent-causal terms, and Steward even argues that understanding free will does not require an account of acting for a reasons in addition to the agent-causal story. But Lowe and O'Connor also separate causality from rationality. All the rational agent can really do, on their account, is to produce something that is still mental and within the sphere of rationality: a volition or an action triggering intentional event. It is then up to that event to account for the actual occurrence of the bodily movement.²⁰² The positing of an entity in between

²⁰¹Even if the willing directly causes a bodily movement, there is nothing in Lowe's account (or in the modified account under consideration here) to rule out non-basic deviance, as in the case of the trampling boards, see §3.3.2.

²⁰²As we saw, Lowe struggles to even give a story about how these volitions produce actual movements, and O'Connor adopts the ordinary event-causal

the rational and the physical is a very natural and understandable move in philosophy. Both Lowe and O'Connor correctly realise that action is not just bodily movement, but also rational movement. Therefore they introduce something that is no longer fully mental—it is “action-triggering” or “the exercise of the will”—but that is at the same time not yet so unlike the mental that it can no longer be rational—it is “intentional” or “rational”. No matter how natural and understandable it is to postulate something in between mind and movement, it is also problematic because any intermediary between the mental and the physical separates them. And such separation invites accidentality to creep in. As Wittgenstein's paraphrase of Schopenhauer with which I started this chapter states: “willing cannot be allowed to stop anywhere short of the action.” As long as agent-causal theories continue to separate the causal production of action from the rationality of action, the two will fall short of one another, and agent-causalism will fail.

4.3 Concluding remarks

After ruling out event-causal accounts of acting for a reason in the previous chapter, this chapter investigated whether contemporary non-causal and agent-causal theories can account for agential control. The non-causal accounts discussed in section 4.1 all failed to rule out accidentality. More importantly, they all failed for the same reason: they had no story whatsoever about the actual material happening of actions. And they cannot tell such a story because they precisely are non-causal accounts.

The agent-causal accounts discussed in section 4.2 also fell short of providing a satisfactory theory of agential control. Their failure is due to the fact that they all try to account for the caus-

account.

ation of action in separation of the rationality of action.

But is such a separation essential to agent-causalism? The event-causal theory of action is a reductive theory of intentional action, it tries to understand the rationality and the occurrence of action in terms of the separate and individually understandable concepts of representation and causation. Therefore the separation of rationality and material occurrence of action is part and parcel of that theory. Similarly, the non-causalist has no way of accounting for the material happening of actions, since it is an indispensable part of that theory that actions can be accounted for non-causally. But agent-causalism does not have the ambition to explain action either reductive or non-causally. Therefore, I believe that it might be possible to construct an agent-causal theory that does not explain the rationality and physicality of the action in separate terms. To develop such a theory we need two crucial ingredients. First we need to understand how rationality can extend all the way into the action. In other words, we need an account of how something can be both fully rational as well as fully material. We want to understand how an action can be “a movement that *is* a thought”, as Sebastian Rödl poetically put it.²⁰³ This will be the aim of the next chapter.

In order to make rational bodily movements less mysterious than they might initially sound, we need to understand they can exist in the natural world. Something I will attempt to explain in Chapter 6. At this point, however, it can seem an impossible task to explain how rational bodily movement can come to be. For we cannot make the causation of intentional action a unique phenomenon, since then we would be resorting to the ‘panicky metaphysics’ of the early agent-causalists. Nor can we account for the causation of actions in separation of their rationality, since then we would make the same mistake as the contemporary agent-

²⁰³See Rödl [2007, p. 18].

causalists.

Nevertheless, I will try to show that it is possible to defend an agent-causal account that does not need a separate supplementary account of acting for a reason and that does not make agent-causation *sui generis* either. I will hold that reasons play an indispensable role in the causation of intentional bodily movements, and at the same time defend that there is nothing *overly* special about the causal production of action. If my account succeeds, we will finally have the contours of a proper understanding of agential control, and therefore there still is hope for the simple picture of free will.

Chapter 5

Practical Knowledge*

*So one might say: voluntary movement
is marked by the absence of surprise.*

—Ludwig Wittgenstein²⁰⁴

When an intentional action is performed something changes in the world. A hand moves, a word is written, a thought expressed. The action *is* this movement, the writing of that word, the expression of this thought. But at the same time the action also has a place in the space of reasons: it is performed in order to attain an end that the agent has in mind. Philosophy of action asks how this is possible, if it is possible at all. How can something be both material and rational, a “movement that *is* a thought”? I believe that this question finds an answer in Elizabeth Anscombe’s book *Intention*.²⁰⁵ Anscombe’s monograph is widely regarded as one of the most important treatments of intentional

*Large portions of this chapter have been drawn from [van Miltenburg 2012] and from several versions of an unpublished paper I am writing together with Dawa Ometto entitled ‘The Factivity of Practical Knowledge’ that will hopefully come out some day soon. I am very grateful to Dawa for allowing me to include text that is partly his.

²⁰⁴Wittgenstein [1953, §628].

²⁰⁵Anscombe [1957].

action and many of the ideas which are now commonplace in the philosophy of action were first expressed in it. However, the main thought expressed in the book—that intentional action is essentially characterised by the agent’s practical knowledge of what she is doing—has been left alone for a long time. One reason for this undoubtedly is the success of Davidson’s causal action theory defended in ‘Actions, Reasons and Causes’ [Davidson 1963] which was published only six years after *Intention*.²⁰⁶ Another reason lies with Anscombe’s book itself. It is written in a style that is “like the confection *panforte*, all fruit and nuts and no dough, very chewy and tough.”²⁰⁷ What further complicates the understanding of her book is that Anscombe herself often seems to be thinking on her feet and is still developing and adapting her theses in the course of the work. All this may lead some to think that it is impossible to find *one* coherent conception of practical knowledge in Anscombe’s book.²⁰⁸ But recently a renewed interest for Anscombe’s thoughts has emerged.²⁰⁹

In this chapter I will explain, develop and defend the thought that practical knowledge is essential to human agency against the background of the contemporary discussion. In the first two sections (§5.1 and §5.2) I will stick closely to Anscombe’s original account and try to explain her views. We will see that Anscombe’s view promises to finally deliver an account of agential control that rules out accidentality and can therefore form the basis for an understanding of the simple picture of free will. However, Anscombe’s view is not without problems. One issue is a problem Sarah Paul calls “deviant formal causation”. Section 5.3 will show

²⁰⁶That theory is explained and criticised in Chapter 3.

²⁰⁷As Anscombe’s daughter puts it in the introduction to M. Geach and Gormally [2005].

²⁰⁸Anselm Müller [1999], for instance, suggests that there are at least two different concepts of practical knowledge at play in *Intention*.

²⁰⁹See, e.g., Falvey [2000], Vogler [2001], Moran [2004], Rödl [2007], Setiya [2008], Teichmann [2008], Thompson [2008], Ford *et al.* [2011], Schwenkler [2011], Horst [2012] and Lavin [2013].

how the Anscombean can avoid it. The next section (§5.4) introduces a much more serious issue for the account of agency in terms of practical knowledge: ‘the problem of failed action’. In section 5.5 I discuss Anscombe’s own treatment of this problem and argue that her solution is largely misunderstood in the contemporary debate about practical knowledge. A proper understanding of her solution will give rise to the idea that an agent’s practical knowledge is *constitutive* of her intentional action—a thought that will be developed and defended in section 5.6.

5.1 The question ‘Why?’

At the start of *Intention* Anscombe distinguishes intentional actions from unintentional behaviour by claiming that the former are “actions to which a certain sense of the question ‘Why?’ is given application” [Anscombe 1957, §5]. The “certain sense” Anscombe has in mind is that the ‘Why?’ question should be answered by giving a reason for action. “Why did you turn on the light?” “Because I want to read *A Game of Thrones*.” But how are we to distinguish this kind of ‘Why?’ question from other kinds? Someone who unintentionally turns on the light can answer the question “Why did you turn on the light?” by specifying a cause: “I was accidentally leaning against the switch”. It would, of course, be hopelessly circular to distinguish the two questions by claiming that the question that asks for reasons is only applicable to intentional movements. Hence, the fact that agents can typically give reasons for their intentional movements by answering ‘Why?’ questions does not by itself provide a helpful criterion of intentionality.²¹⁰ Nevertheless, in order to find such a criterion it is

²¹⁰At best, the agent’s capacity to answer ‘Why?’ questions provides a behaviouristic criterion of which movements are intentional and which are not. Moran and Stone [2011] extensively discuss a behaviouristic reading of Anscombe and show why it should be rejected. Anscombe herself attacks behaviourism in [Anscombe 2005].

useful to consider when the ‘Why?’ question that asks for reasons is refused application.

The question “Why are you φ ’ing” does not apply when the agent is altogether unaware that she is φ ’ing. Consider someone who is leaning against a doorpost and is asked “‘Why are you ringing that bell’ [and] replies ‘Good heavens I didn’t know *I* was ringing it!’” [Anscombe 1957, §28]. That the question ‘Why?’ does not apply in such cases is very clear—if the agent does not even know that she is φ ’ing she obviously cannot say *why* she is φ ’ing either.²¹¹ Hence, the least that is required for an action to be intentional is that the agent knows that she is performing that action.

A more interesting and less obvious case where the ‘Why?’ question does not apply is when an agent’s awareness of what she is doing is merely observational. Charlie, for instance, who is moving her leg but only knows this because she sees its movement and feels that the doctor strikes her knee with a reflex hammer, can give a cause of, but no reason for her action. Agents who act for a reason do not need to check and see what they are doing. They know it “without observation” [ibid, §8]. Consider Elliot who is cutting some onions in order to make risotto. Any observer can see that she is cutting onions, but by observing these movements the observer cannot yet determine whether Elliot is cooking risotto or, say, spaghetti. To find that out through observation, the observer has to wait and see what she does next. Perhaps Elliot’s intentions

²¹¹If we had described the agent’s movements differently, application of the ‘Why?’ question might not be refused. ‘Why are you leaning against the doorpost?’ can have an answer: ‘To rest a bit’. Thus whether an agent can give reasons for certain of her movements depends on how these movements are described. This is why Anscombe claims that actions are only intentional *under a description*. I may intentionally put a book on the table, but with the same movements unintentionally put the book on a puddle of ink. See Anscombe [1979]. The thought that actions are intentional under a description is also accepted by Davidson and has become commonplace in contemporary philosophy of action.

become clear later when she takes rice instead of pasta out of the cupboard. How different is the situation for Elliot herself. She already knew from the moment she made up her mind and walked over to the kitchen that she was cooking risotto. Elliot does not have to find out what she is doing by observing her own movements. Her knowledge is “in intention” [Anscombe 1957, §32]: she knows that she is cooking risotto because that is what she intends to do.²¹²

That agents have such non-observational knowledge of their own intentional actions, according to Anscombe, is no mere eccentricity. To the contrary, it is the defining feature of acting for a reason.²¹³ Intentional action essentially is action of which an agent has a special kind of non-observational knowledge: *practical knowledge*. It is important to note that practical knowledge is only a *species* of non-observational knowledge. Anscombe, for instance, mentions the knowledge we have of the position of our own limbs as another kind of non-observational knowledge. The latter knowledge, however, is not practical in that it is not knowledge of our intentional action—I can know where my arm is without currently intending to move that arm in any way. It is perhaps an unfortunate fact in the history of contemporary action theory, that Anscombe introduced the concept of practical knowledge, as a kind of knowledge that is acquired without observation, for it has led many commentators on her work to focus only on that feature.²¹⁴ But the fact that practical knowledge is non-observational

²¹²Cf. Anscombe [1957, §28]: “ Say I go over to the window and open it. Someone who hears me moving calls out: What are you doing making that noise? I reply ‘Opening the window’ [...] But I don’t say the words like this: ‘Let me see, what is this body bringing about? Ah yes! the opening of the window’.”

²¹³If Anscombe is right that intention is essentially conscious—and this chapter argues that she is right—then the entire neuroscientific discussion of ‘unconscious intentions’ we encountered in Chapter 1 seems even more far-far-ginious than it already did.

²¹⁴Falvey [2000], Pickard [2004] and Paul [2009] for instance, are all solely concerned with the thought that the knowledge an agent has of her own ac-

merely provides a negative characterisation of it, it only tells us one thing that practical knowledge is not.

We can arrive at a positive understanding of practical knowledge by considering the question ‘Why?’ again. That question asks for reasons. And indeed an agent typically not just is non-observationally aware of *what* she is doing, she is also able to say *why* she is doing it. Consider the series of ‘Why?’ questions in Anscombe’s [1957, §23] example of a gardener pumping poisoned water:

- ‘Why are you moving your arms?’ (A) – ‘I’m pumping water’ (B)
- ‘Why are you pumping water?’ (B) – ‘I’m replenishing the water supply’ (C)
- ‘Why are you replenishing the water supply?’ (C) – ‘I’m poisoning the inhabitants’ (D)

Each answer to the ‘Why?’ question has the following form: ‘I am *A*-ing because I am *B*-ing’.²¹⁵ Each answer thus provides a reason for acting by re-describing the gardener’s act in a particular way. We come to understand the rationality of the arm movements once we understand that they constitute a pumping. But since ‘pumping’ itself is also a description of the gardener’s intentional action the question ‘Why?’ again applies to it: we can ask why the gardener is pumping. He could then reply with another re-description, about which we can again ask another ‘Why?’ question, etc. In this way a series of ‘Why?’ questions reveals the

tions is non-observational. John McDowell [2011] even explicitly attempts to understand agential knowledge of intentional actions in terms of the knowledge an agent has of the position of her own limbs. As we will see in section 5.3 the failure to distinguish practical from other kinds of non-observational knowledge can also lead one to confuse the agent’s aims with the foreseen side-effects of her action.

²¹⁵See part two of [Thompson 2008] for an argument that action explanation is fundamentally of this form.

rational structure of *means* and *ends* contained within the action. Hence, practical knowledge is not merely the knowledge that one is doing something under one particular description. Rather, it is knowledge of all the descriptions under which the act is intentional, and the way they stand in means-end relations to each other. It is knowledge of the rational structure of one’s action.

The rational structure of action that is brought out by the ‘Why?’ question is mirrored in the agent’s *practical reasoning*. ‘How can I poison the inhabitants?’ ‘By replenishing the water supply.’ ‘How can I replenish the water supply?’ ‘By pumping water’. In practical reasoning this deduction of means from ends continues until the agent arrives at something she can directly engage in, e.g., moving her arms.²¹⁶

In the example above, the descriptions of the gardener’s actions are descriptions of the same movement. But, as Michael Thompson [2008] notes, this is not essential to Anscombe’s account. Take Elliot who we left chopping onions in the kitchen. To explain why she is cutting onions, she can re-describe her movements as risotto making. But making risotto involves much more than just onion cutting. Ingredients have to be taken out of the fridge, the stove needs to be turned on, garlic has to be peeled, chopped and fried, wine needs to be added, etc. Without Elliot’s practical knowledge, all these actions would be nothing more than a bunch of random movements in the kitchen. It is Elliot’s knowledge about what she is doing that *unifies* all these movements into one action: risotto cooking. The means-end order known practically by the agent is not something that is out there and ready to be grasped by the agent. To the contrary, it is the agent’s practical knowledge that brings about the existence of that order. Elliot only cuts onions because she knows that she is cooking risotto.

²¹⁶Anscombe offers an extensive account of practical reasoning in her “Von Wright on practical inference” [1974]. For an elaborate discussion of her views see Müller [1979] and Rödl [2011].

The gardener only pumps because he knows that this is a means for replenishing the water supply.

[T]he account given by Aquinas of the nature of practical knowledge holds: Practical knowledge is ‘the cause of what it understands’, unlike ‘speculative’ knowledge, which ‘is derived from the objects known’. [Anscombe 1957, §48]

What does it mean to say that something is the cause of what it understands? Consider that I can have ordinary (speculative) knowledge about my cat. I know that she is black for instance. I know this because I saw that she is black. In this way, the object (my cat) caused my belief that it (she) is black. How different is the knowledge we have of our own actions. My knowledge that I am writing this section is not caused by the section—right now I know that I am writing it even though it does not yet (fully) exist. It rather is my knowledge that I am writing this section that makes me move my fingers in specific ways. Intentional actions are caused by our knowledge of them. It is this feature, according to Anscombe, that ultimately distinguishes intentional actions from mere behaviour.

5.2 The cause of what it understands

At the end of the last section we saw that Anscombe does think that practical knowledge play a *causal* role in the production of intentional action. In that regard her account differs from the non-causal accounts defended by her fellow Wittgensteinians.²¹⁷ But her account is also importantly different from the standard (event-)causal theory of action discussed in Chapter 3. On that view, whether an event is an intentional action depends solely on

²¹⁷Although this is sometimes neglected. Davidson, for instance, includes her among the non-causalists that are the target of his argument that reasons are causes. See Davidson [1963, fn. 1].

the causal relations it has to antecedent events with the right content. For Davidson there is no *intrinsic* difference between an arm raising and an arm rising. Arm raising only is intentional because it is caused by other events that represent it. Anscombe's view is different. Practical knowledge is not "a mere extra feature of events whose description would otherwise be the same" because "without it what happens does not come under the description—execution of intentions—whose characteristics we have been investigating" [Anscombe 1957, §48]. In this way, practical knowledge determines the *form* of what happens: it provides actions with their rational structure of means and ends. Elliot's practical knowledge that she is making risotto, determines that the description "cooking risotto" applies to her onion cutting. In other words, Elliot's practical knowledge determines that her onion cutting *is* risotto cooking. Or, to use a bit of Aristotelean jargon, we can say that practical knowledge is the *formal cause* of the action it understands.²¹⁸

At this point we can begin to see why Anscombe's account might not fall prey to the deviant causal chain scenario that so plagued Davidson's theory. The climber who accidentally lets go of the rope, does not have practical knowledge that he is loosening his grip. He knows only by observations that his hands are trembling so vigorously that he no longer can hold on. Furthermore, he also does not think of these movements as having a place in an order of means and ends. Therefore, it seems that Anscombe can simply

²¹⁸Although Anscombe herself does not explicitly appeal to the concept of *formal causation*, her theory is often described in terms of it. See, for example, Moran [2004], Teichmann [2008] and Rödl [2011]. The formal cause is one of the four causes Aristotle distinguishes—see *Physics* [Aristotle 1996], Book II.5, and *Metaphysics* [Aristotle 1998], Book V.2. However, those who use of this label in the contemporary discussion do not aim to fully conform to Aristotle's views—they rather use the term to distinguish the way in which practical knowledge causes an action to be the action that it is, from the way in which action is caused by mental states on the causal theory of action. To the latter sort of causation we could refer as 'efficient causation' to use another of Aristotle's labels.

deny that the climber's loosening of his grip is intentional—it lacks the rational structure essential to intentional agency.²¹⁹ But, as I have argued in section 3.4, this problem of causal deviance is only a symptom of a general failure to rule out accidentality. Does Anscombe's account only treat some symptoms or is it able to cure the disease?

If practical knowledge merely causes the form of movements without influencing their material production, then it seems that it is no more than a lucky accident that the movements correspond to the practical thought of the agent.²²⁰ Anscombe writes:

I once saw some notes on a lecture of Wittgenstein in which he imagined some leaves blown about by the wind and saying 'Now I'll go this way . . . now I'll go that way' as the wind blew them. The analogy is unsatisfactory in apparently assigning no role to these predictions other than that of an unnecessary accompaniment to the movements of the leaves. But it might be replied: what do you mean by an 'unnecessary' accompaniment? If you mean one in the absence of which the movements of the leaves would have been just the same, the analogy [with intentional action] is certainly bad. [Anscombe 1957, §3]

The problem with the analogy between the thoughts of the leaves and practical thought, Anscombe says, "is not that it assigns a false role to our intentions, but only that it does not describe their role at all." [ibid, §3]. If the leaves were agents then their thoughts would not be idle, their thoughts would *produce* their movements. Consider Elliot again, she only starts cutting onions because she knows that this is a means to risotto making. Without practical knowledge she would simply have no reason at all to engage in onion cutting. Her practical knowledge of what she is doing is no further feature that exists in addition to her movements, but some-

²¹⁹See Vogler [2001] for a more elaborate explanation of how Anscombeans can deal with the problem of causal deviance.

²²⁰John Schwenkler [forthcoming] meticulously points this out.

thing without which her movements would not be happening at all. Hence practical knowledge does not merely determine the descriptions under which Elliot's movements are intentional, it also produces the movements themselves.²²¹ It is however, important to notice that the way in which practical thought produces movements is radically different from the way in which intentions cause what they represent on Davidson's account. On that account an intention is a mental event that stands in an extrinsic, causal relation to the agent's movement. But on Anscombe's account the agent only moves because she herself understands what role her movements play in the rational structure of the action she is performing. In this way the agent's movements are truly caused in virtue of the content of her thought via her recognition of this content.²²² As we will see below (§5.6), the agent's practical thoughts *constitute* her movement.

The best way to understand the productive force of practical knowledge, I believe, is to understand it as an intrinsically rational *power*. Indeed, at several points in *Intention*, Anscombe herself speaks about practical knowledge as a capacity that can be exercised:

When we ordinarily speak of practical knowledge [in the sense of know-how] we have in mind a certain sort of general capacity in a particular field [...] In the case of [general] practical knowledge the exercise of the capacity is nothing but the doing or supervising of the operations of which a man has practical knowledge; but this is not just the coming about of certain effects, [...] for what he effects is formally characterised as subject to our question 'Why?' whose application displays the A-D order which we discovered. [Anscombe 1957, §48]

²²¹The productive role of practical knowledge becomes even more apparent if we consider cases in which it is suddenly absent. David Velleman [1989, p. 15] offers such an example: an agent is walking down Fifth Avenue, forgets what she is doing and stops walking. It is only after she remembers that she was, say, buying some groceries that she is able to continue on her way.

²²²Cf. §3.3.3.

The exact nature of this rational power of practical knowledge, and the question how there can be such a power, will be a recurring topic in this chapter and the next. For now I want to focus on the nature of the effects this power produces. In the previous chapter (§4.2) we encountered agent-causal accounts that also construe acting in terms of powers. These accounts argue that agents have the power to bring about volitions or intentions that subsequently lead to action. But for Anscombe the effect of the agent-causal power of practical knowledge is not something that eventually leads to action. What the agent effects is “formally characterised as the subject of our question ‘Why?’”, in other words, it is the intentional action itself. For Anscombe, there is no gap between the exercise of practical knowledge and the material happening of our intentional actions. As she puts it: “I *do* what *happens*” [Anscombe 1957, §29]. Anscombe does not need an intermediate mental state that somehow transfers its rationality to essentially material movements, because she considers actions themselves to be intrinsically rational. As Michael Thompson explains, this thought is the ‘overarching thesis’ of Anscombe’s account:

[Practical knowledge] extends beyond the inner recesses of the mind, beyond the narrowly psychical, and into the things that I am doing. [Thompson 2011, p. 200]

This acceptance of intrinsically rational material happenings is, to say the least, controversial. But I believe that we must accept such irreducible teleology in order to understand intentional action. As we saw in Chapter 3 many have tried to reduce intentionality, but no one has succeeded.²²³

If Anscombe is right that our practical knowledge does not stop short of the facts, that we *do* what *happens*, then nothing can come between the agent’s practical thought and her movements.

²²³Much more on the possibility of rational happenings in nature will follow in the next chapter (§6.5).

Hence it is impossible that the correspondence between practical thought and movement is merely accidental. Without the possibility of accidentality, actions are not up to luck: the agent herself can truly be in control of what she does. Thus Anscombe's account promises to deliver an account of agential control that can finally help us to maintain the simple picture of free will. Unfortunately it is precisely the thought that there is no gap between thought and movement that many contemporary philosophers criticise or fail to understand. And, as we will see, even many so-called neo-Anscombeans reject exactly this part of her theory. In the remainder of this chapter I will look at two different criticisms of Anscombe's action theory and show how an Anscombean can try to evade them. In answering these criticisms much more will become clear about the nature of practical knowledge and the happenings it produces.

5.3 Practical knowledge and side-effects

Anscombe focuses on the knowledge agents have of their intentional doings. But an agent might also know that her movement leads to certain effects that she does not intend to happen. I might intend to play the trombone and foresee that this could wake the neighbour's baby. Nevertheless there is a difference between me playing the trombone in order to wake the baby, and my playing in order to practice for the upcoming concert of which the baby's waking is a mere side-effect. According to Sarah Paul [2011], Anscombeans cannot make this distinction between aim and foreseen side-effect.

Paul sketches two scenarios in which different gardeners go through the same movements of pumping to replenish a house's water supply with poisoned water. However, the two gardeners have different purposes. The first acts with murderous intent, while the second is indifferent about the potentially deadly side-

effects of his action and merely wants to earn his pay. Hence, even though both gardeners make the same movements it would be wrong to say that both Murderous and Indifferent Gardener perform the same intentional actions. But how are we to distinguish their actions? Since the two scenarios are set up in a way that holds fixed many of the possible different grounds for a distinction (equal contexts, the same physical movements), “it seems that the asymmetry can only be traced to the divergent psychological properties of the two gardeners” [Paul 2011, p. 11].

On Anscombe’s view the agent’s knowledge is the determinant of what she is intentionally doing. Therefore the Anscombean has to spell out the asymmetry in terms of a difference in both gardeners’ knowledge. According to Paul, this is impossible because both gardeners possess the exact same knowledge. Both knew beforehand that the water was poisoned and that the inhabitants would drink water from their cistern. Therefore both gardeners know that they are poisoning the inhabitants. And, Paul stresses, both know this non-observationally—neither has to look in order to find out that he is replenishing the cistern with poisoned water. Paul concludes that on Anscombe’s account too much of what is happening comes out as something we intentionally do. The indifferent gardener is not intentionally poisoning despite his knowledge that this is a consequence of him earning his pay. Hence it seems that there must be a gap between the things we know we do and the things we do intentionally—we simply know that we do certain things without intending them.

Paul dubs this problem for the Anscombean “the problem of deviant formal causation”. The agent’s knowledge of the side-effects of her intentional action seems to deviantly turn these side-effects into parts of the action itself. This problem mirrors the problem of causal deviance for the standard event-causal theory of action. On Davidson’s view the climber’s loosening of his grip

wrongly comes out as intentional, on Anscombe's view Indifferent Gardener's poisoning befalls the same fate.

However, the problem for the Anscombean is quite easy to overcome. Anscombe claims that an agent's actions are only intentional under the descriptions of which she has *practical* knowledge. But it is important to note that not just any knowledge of action is practical knowledge. One feature of practical knowledge is that it is non-observational. If this was all there is to it, then Paul would surely be right that practical knowledge cannot be the defining feature of intentional action because we also non-observationally know some of the unintended side-effects of our actions. But, as we have seen, practical knowledge is not just non-observational, it also is *the cause of what it understands*. Non-observational knowledge of action is practical only when it plays a specific role in the coming about of the action that it represents. Even though both gardeners might know the same things non-observationally, the two gardeners might not have the same practical knowledge. However Paul thinks otherwise:

If it is necessary and sufficient for the formal causation of an intentional poisoning that the agent knows he is bringing about a poisoning in virtue of the practical efficacy of that very bit of knowledge, the poisoning comes out as intended in the case of the Indifferent Gardener as well as the Murderous Gardener.

[Paul 2011, p. 12]

Thus, Paul claims that the knowledge of both Murderous and Indifferent Gardener is equally practically efficacious. But this is simply false. Indifferent Gardener's knowledge that he is poisoning is not efficacious. If Indifferent Gardener had been ignorant of the fact that the water was poisoned, and thus had not known that his replenishing would bring about the poisoning, he would still have replenished the water supply. Replenishing, after all, is a fine means for earning his salary. Compare this with Murderous

Gardener. *He* would not have replenished if he had not known that it was a means to poisoning, since the *point* of his replenishing consists only in killing the inhabitants. Hence Murderous Gardener's knowledge that he is poisoning clearly *is* practically efficacious, while Indifferent Gardener's knowledge is not.²²⁴ Only Murderous Gardener's knowledge assigns poisoning a role in the structure of means and ends intrinsic to his action. Therefore, only Murderous Gardener's movements constitute an intentional poisoning. Indifferent Gardener's movements do not, because his (non-observational) knowledge that the replenishing brings about a poisoning is not practically efficacious. For him, the poisoning is merely a foreseen side-effect.

Once we realise that non-observational knowledge is a *genus*, of which practical knowledge is but one *species*, it is clear that the Anscombean is, after all, able to make the distinction between aim and foreseen side-effect. Aims are known practically, whereas an agent can only have speculative knowledge of side-effects. Indeed, knowledge of side-effects is derived from the object known. Indifferent Gardener derives the knowledge that the inhabitants will be poisoned from his practical knowledge that he is replenishing and his prior speculative knowledge that the water is poisoned and that the inhabitants drink from the cistern. Only Murderous Gardener's knowledge causes its object. He would add poison if the solution was not strong enough and he would try to persuade the inhabitants to drink from the cistern if he noticed that they might not do so.

However, Paul objects that it is *ad hoc* to deny that Indifferent Gardeners knowledge that he is poisoning the inhabitants is practical. According to her it *is* practical because it should play

²²⁴To be clear, I am not claiming that Murderous Gardener's thought is efficacious and Indifferent's is not because these counterfactuals hold. Rather, these counterfactuals hold because only Murderous Gardener's thought is practical, and therefore productive. The precise role practical knowledge plays in the production of action will become clearer below (§5.6).

a role in his practical reasoning. After all, she argues, Indifferent Gardener knew full well that the water was poisoned and therefore made his decision “in the light of his expectation that it will result in the poisoning [...] and a reconstruction of how he reached that decision ought to include that expectation” [Paul 2011, p. 14]. But this presupposes a certain picture of practical reasoning Anscombe does not endorse. On Anscombe’s view, practical reasoning is the deduction of means from ends. Indifferent Gardener wants to earn his pay which he thinks he can achieve by replenishing the water supply, which he in turn can do by operating the pump, etc. The fact that the water is poisoned does not figure in this deduction at all. Actually, the Anscombean could hold that the indifference of the gardener precisely consists in his failure to include his knowledge that his action will result in a poisoning into his practical reasoning.²²⁵

We can conclude that the Anscombean is not only able to distinguish aim and side-effect, she is able to do so in a way that meshes nicely with the overall project of analysing the intentionality of actions in terms of the sort of knowledge about the action possessed by the agent. We saw in the previous sections that what the agent does is intentional under the descriptions of which an agent has practical knowledge and unintentional or merely accidental under descriptions the agent is unaware of, or merely aware of observationally. We can now extend this taxonomy with another category: the category of descriptions the agent is non-observationally aware of but of which she does not have practical knowledge. In fact Anscombe herself already recognises this category.²²⁶ And she calls actions that fall in it: “voluntary though

²²⁵ Another objection Sarah Paul raises against separating aim and foreseen side-effect in terms of knowledge is that knowledge necessarily has the wrong *direction of fit* for this job. Much more on that notion will follow in §5.5.

²²⁶ She even has famously employed the distinction between intended aim and side-effect in her defence of the doctrine of ‘double effect’, i.e., the idea that it can be morally permissible to cause harm as a side-effect (or double effect) of

not intentional”:

Something is voluntary though not intentional if it is the antecedently known concomitant result of one’s intentional action so that one could have prevented it if one would have given up the action; but it is not intentional: one rejects the question ‘Why?’ in its connexion. [Anscombe 1957, §49]²²⁷

In the contemporary free will debate, free action is thought to be a subclass of intentional action. But for Anscombe it is the other way around. The class of free (or at least voluntary) actions is larger than the class of intentional actions.²²⁸ However, on Anscombe’s view, the freedom of voluntary actions seems—at least for human agents—to be grounded in intentional action. As Anscombe says, the side-effects of one’s intentional action are voluntary because we can prevent them by giving up the action. But giving up an action is something we do intentionally. If I want to prevent waking my neighbour’s baby, that becomes my aim, and one means to achieve this aim is to refrain from playing the trombone right now. Hence we must continue our quest to understand intentionality in order to understand freedom. Let us therefore turn to a second and much more serious problem for Anscombe’s idea that there is no gap between what we intentionally do and what happens.²²⁹

doing something good, even though it would not be permissible to cause harm as a means to the same good end. See Anscombe [1982].

²²⁷According to Anscombe the category of the voluntary includes more than just foreseen side-effects. Actions are also voluntary if one does not reject the question why, but gives answers like ‘I was fiddling’, ‘it was a casual movement’ or ‘I don’t know why’. Furthermore, movements are voluntary when they are not considered by the agent though she can say what they are when she considers them. Finally, all intentional actions are also voluntary.

²²⁸Only Helen Steward seems to diverge from the contemporary consensus. As we saw (§4.2.1) she thinks that all animal self-movement—and not merely intentional self-movement—is free in the sense that it requires indeterminism. Cf. the quotes passage on 153.

²²⁹Perhaps the actions of animals are also voluntary in the sense that they can stop their actions when they recognise that they lead to unwanted side-effects. Therefore, Helen Steward may well be right that animals also are the

5.4 The problem of failed action

Anscombe complains in *Intention* that everyone who hears her formula “I *do* what *happens*” finds it “extremely paradoxical and obscure” [Anscombe 1957, §29]. Unfortunately the same is still true today. Many authors insist that we can simply be wrong about what we are doing and this can cause a gap between what we think we are doing and what is in fact happening.²³⁰ Considering an agent who writes something down without looking at his paper, Moran writes:

The agent can, of course, be flatly wrong about what he takes himself to be doing, for instance [...] when the pen runs out of ink without his noticing, or more generally when the empirical conditions enabling a particular action fail to obtain. If he is wrong in assuming that writing is getting produced, then he cannot have practical knowledge that he is writing. [Moran 2004, p. 60]

The writer surely is *intending* to write, but in this case it seems that his intention does not deliver him any knowledge about what is happening. To find out whether he really is writing he has to check and see whether any words are getting produced. Given the possibility of failure in the performance of the action, it may seem that an agent can never be sure that she is doing what she intends to do without observation. A sceptic might wonder whether we can ever possess practical knowledge about what we are doing. Hence, cases of failed action seem to threaten Anscombe’s main idea that practical thought extends beyond the inner recesses of the mind and is knowledge of what is happening.

“settlers of matters” [Steward 2012a, p. 55]. But, it seems to me that animal actions are not intentional. At least not in the sense that animal thought is the cause of what it understands. For, I think that an animal does not have to recognise and understand the purpose of its action in order to engage in it. Developing this suggestion further would, however, go far beyond the scope of the current text.

²³⁰See, e.g., Falvey [2000], Grünbaum [2009] and Schwenkler [forthcoming].

Some contemporary writers think that we can easily dismiss this worry. They insist that the possibility of error does not exclude the presence of knowledge in case the agent is successful. Such a dismissive response to the problem of failed action has even been attributed to Anscombe herself. David Velleman reflects on Anscombe's statement that "the failure to execute intentions is necessarily the rare exception" [Anscombe 1957, §48], and writes:

Anscombe was also, as I interpret her, a reliabilist about knowledge—in particular, about what is "known by being the content of intention". She thought that a reliable connection in general between what's intended and what's done is sufficient to confer the status of knowledge on a particular intention, provided that the connection holds up in the particular case. [Velleman 2004, p. 227]²³¹

Whether or not this is a correct interpretation of Anscombe (which I think it is not), the dismissive strategy fails to appreciate the true threat posed by the idea that an agent may be 'flatly wrong' about what she takes herself to be doing. For, it is not sufficient to merely explain how we can still have knowledge when we act successfully. The Anscombean also has to explain how there can still be intentional activity when an agent fails to successfully execute her intentions. The writer who fails to produce legible words because his pen has run out is still doing something intentionally. But the Anscombean claims that intentionality essentially involves practical knowledge, so how can the writer's hand movements be intentional in the absence of such knowledge?

One suggestion is that although the writer does not know that he is writing (because no words are appearing on paper), he still has practical knowledge that he is moving his hands. And the

²³¹Cf. [Teichmann 2008, p. 25]: "If doubt arises as to whether I am in fact doing what I say I am doing, then I will need to resort to observation, or to asking others. But doubt is the exception, not the rule, and lack of doubt is the default position."

practical knowledge of his hand movements explains their intentionality. As Thor Grünbaum states:

The fact that we have this separation [between what the agent does and what she thinks she does] when the action is conceived under some descriptions does not even in the extreme cases exclude that intentionality and practical knowledge could coincide under a different description. [Grünbaum 2009, p. 45]

But, for the Anscombean, this will not do. Recall that practical knowledge is not just knowledge *that* one is doing something, it is also knowledge of *why* one is doing it. The question ‘Why are you *A*-ing’ typically receives an answer of the form ‘I am *A*-ing because I am *B*-ing’. Hence practical knowledge is not just knowledge of the intentional action under one description. It is knowledge of all the descriptions under which the action is intentional. And since these descriptions are related to each other as descriptions of means and ends, practical knowledge is knowledge of the rational structure contained within action.²³² As Moran points out, this feature of practical knowledge radically distinguishes it from speculative (observational) knowledge. In case speculative knowledge fails, we can indeed often fall back to a ‘weaker description’ of what happened:

I claim that there is a goldfinch at the bottom of the garden, but my companion corrects me. I then may say, ‘Well, it’s either a goldfinch or a goldcrest’, and then perhaps I have to retreat from this too, and so on. [Moran 2004, p. 64]

But in case any of the links in the means-ends chain fails to be practically known by the agent, we cannot retreat to such a weaker description. Consider our writer again. No words are appearing on paper so he does not have practical knowledge of his action under the description ‘I am writing’. Grünbaum suggests that this is unproblematic, since the writer might still have practical

²³²See §5.1 above.

knowledge that he is moving his hands. But what if we were to ask the writer why he was moving his hands? He could not reply, “I am moving my hands because I am writing” since it is false that he is writing. Therefore it seems that the question ‘Why?’ that asks for reasons does not apply to the writer’s hand movements. However, on the Anscombean account, an action is intentional only if that question ‘Why?’ applies to it.²³³ Hence, the writer’s hand movements do not come out as intentional after all.

This shows that a single error in the agent’s practical thought can bring down the entire means-end chain contained within his action. In the case of practical knowledge there simply is no ‘weaker’ description of her action that the agent still conceives of as the point of what she is doing. The writer would not have moved his hands if he had not thought that this would produce any writing.

Failed actions, then, are not an oddity the Anscombean can safely ignore—even if they are “necessarily the rare exception”. Whenever there is a failure in the execution of an intention, we are left without the possibility of giving an intentional explanation of the agent’s movements under *any* description—including the more immediate ones. Yet even if the agent is unwittingly failing to execute her intention, there is still clearly intentional movement going on. It is just *false* to say that the question ‘Why?’ does not apply to any description of what she is doing. As Moran says, “The agent is plainly doing something; it is not the case that these movements lie outside of action description altogether” [Moran 2004, p. 63]. But practical knowledge was meant to *mark off* the sphere of action description. The possibility of the agent’s being flatly wrong about what she takes herself to be doing would thus demonstrate Anscombe’s failure to give a positive characterisation of the sense of the question ‘Why?’, and thereby her failure to arrive at a cogent account of intentional action altogether.

²³³As we saw above (§5.1).

Hence, the Anscombean cannot retreat to more immediate descriptions of what happens in cases of failure of practical knowledge. But is there no other way out? Some contemporary philosophers inspired by Anscombe would wish to point out that, even if it is not true that the agent is writing, we can still give an intentional explanation of her hand movements. Can an agent not simply answer the question ‘Why are you moving your hands?’ by saying: ‘I *intend* to write’? That the agent intends this seems true no matter whether the pen has run out of ink or there is any other sort of interference. And given a suitable understanding of the state of intention, this proposal might even salvage a sense in which practical knowledge is essential to a positive characterisation of intentional action. For if intentions are *beliefs* that one is performing a certain action, then in any case of a *successful* execution of an intention, the belief will be true, and the agent will have knowledge of what she does. Here the analogy with speculative knowledge survives: just as we are sometimes forced to say that we did not *know* that there is a gold finch at the bottom of the garden, but merely believed it, we can sometimes fall back to the claim that we *intended* to replenish the water supply, although we were wrong in thinking that we were really doing it.²³⁴

But this avenue of escape departs from the Anscombean project in an important respect. It assumes that, when an agent answers the question ‘Why?’ with something like ‘I’m writing’, that reply derives its rationalising potential from the underlying *intention* to write. Whether or not she is really writing is irrelevant to the explanation of the intentionality of her hand movements. So whether the kind of explanation demanded by the question ‘Why?’ applies is *independent* of the agent’s thought being knowledge. However, the problem posed by failed action arose precisely because the agent’s possession of practical knowledge was

²³⁴The main exponents of this intention-as-belief approach are Velleman, e.g., [1989] and Setiya, e.g., [2011].

not seen as accidental to the application of the question ‘Why?’. The intention-as-belief approach thus avoids the problem only at the cost of assuming that we can understand the sense of the question ‘Why?’ without reference to practical knowledge. Nor could the approach elucidate, as Anscombe does, the sense of the question by reference to the means-ends relation in which its answers stand: for one’s means to, e.g., poisoning the inhabitants is not being in the *state* of intending to replenish, but the intentional *act* of replenishing. As Moran & Stone [2011] correctly observe, the idea that an answer to the question ‘Why?’ must itself fit into the means-ends chain is one of Anscombe’s main insights. They argue that we must accordingly understand an agent’s intending to ϕ not as a state *preceding* the agent’s action (as the intention-as-belief approach must understand it), but as already located on the “spectrum of unfolding action” [ibid, p. 49], i.e., as *part* of her intentional activity.

On the intention-as-belief picture we would need a different understanding of the relation between an agent’s state of intending to ϕ and her actual intentional movements. The most obvious solution to this is to view the intentional state as the *mental cause* of the action.²³⁵ But this takes us back to the causal theory of action, and as I argued in Chapter 3, there are very good grounds for rejecting that theory. Anscombe’s account of agency in terms of practical knowledge was supposed to offer an alternative to the standard causal theory. And indeed many recent authors argue that her account holds the key to an altogether different understanding of intentional action.²³⁶ Moreover, it is this understanding of intentional action that promises to deliver a cogent account of the control required for free will. Hence if we want to reap the fruits of Anscombe’s action theory for an understanding of free

²³⁵Both Velleman and Setiya are explicitly committed to the project of accounting for practical knowledge in terms of a broadly causal theory of action.

²³⁶See e.g. Rödl [2007], Thompson [2011] and Lavin [2013].

will, this requires us to dismantle the problem of failed action in a way that does not give up the idea that the sense of the question ‘Why?’ depends on the agent’s possession of practical knowledge.

5.5 The mistake is in the performance

Interestingly, Anscombe herself was well aware of the problem of failed action. Indeed, the example of the writer whose pen runs out is drawn from *Intention*. However, Anscombe seems to shrug off the problem by claiming that ‘the mistake is in the performance, not in the judgement’ when an agent fails to correctly execute her intention. Considering a man who intends to write something on a blackboard, Anscombe says:

That intention for example would not have been executed if something had gone wrong with the chalk or the surface, so that the words did not appear. And my knowledge would have been the same even if this had happened [...] Someone might say that it was a funny sort of knowledge that was still knowledge even though what it was knowledge of was not the case! On the other hand Theophrastus’ remark holds good: ‘the mistake is in the performance, not in the judgement’. [Anscombe 1957, §45]²³⁷

This remark that the mistake is in the performance is commonly seen as a reminder of the special *direction of fit* of intention.²³⁸ And indeed, Anscombe’s *Intention* is often considered to be the *locus classicus* of the notion of direction of fit.²³⁹ She considers two lists: one is the shopping list of a man who is doing the gro-

²³⁷Anscombe attributes this phrase is adapted from Theophrastus, *Magna Moralia* 1189b 22, but remarks that it is uncertain whether Theophrastus really was the author of it. Today it is commonly held that Aristotle was the author.

²³⁸See, e.g., Moran [2004] and Haddock [2010].

²³⁹The phrase ‘direction of fit’ does not occur in *Intention*, it was first coined in the context of speech act theory by J.L. Austin [1952]. John Searle brought the notion into the philosophy of mind [Searle 1983]. More on this below (p. 195).

ceries, and the other is a list of a detective who is following the man around, jotting down which items he purchases [Anscombe 1957, §32]. The difference in direction of fit between the two lists consists in the conditions under which each is mistaken: the detective's record is wrong if he writes down items that the man has not purchased, but the shopping list is *not* wrong when the man accidentally comes home with butter instead of margarine. Analogously, when the man thinks 'I am buying margarine' and accidentally brings home butter, it is his *action* that is mistaken, and not his thought—while if the detective falsely believes that the man bought margarine it is his *thought* that is mistaken. It is this difference in direction of fit that distinguishes practical thought from speculative or contemplative thought such as ordinary observational knowledge.

However, it is unclear how an appeal to direction of fit can answer the problem of failed action:

To disqualify as knowledge, it doesn't seem to matter where the error comes from so long as there is error. 'Direction of fit' considerations are not the point here. The distinction between practical and speculative knowledge does not concern the requirement of truth, but the question of what is to be corrected in cases of failure of fit. [Moran 2004, p. 61]

Thus Moran suggests that Anscombe arrives at the paradoxical claim that an agent has practical knowledge even in case something goes wrong because she is mistaken about the implications of the difference between the direction of fit of practical knowledge and speculative knowledge:

[Anscombe] seems to confuse the requirement of truth for knowledge, which applies to any knowledge practical or speculative, with the question of whether it is the action or the putative knowledge that is to be corrected in the case of disparity. [ibid, pp. 60–61]

According to Moran, whether the mistake is in the performance or in the thought, only determines what needs to be altered when a discrepancy between world and thought occurs.²⁴⁰ An agent's practical thought about what she is doing differs from ordinary belief only in that it is the world (one's behaviour) that must be adapted, instead of the thought. Indeed, this view aligns nicely with the contemporary account of the direction of fit of *propositional attitudes*. As Searle [1985, 2001] argued, we can take several different attitudes, e.g., believing, wishing, hoping, desiring etc., towards a proposition *p*. Each attitude comes with its own direction of fit.²⁴¹ If I *believe* that 'van Gogh' is written on this page, I will change my belief if it turns out that 'van Gogh' is not written on this page. If I *desire* that 'van Gogh' is written on this page, and find out that it isn't, I will proceed by writing it. But the proposition 'van Gogh' is written on this page' is true or false, independently of my attitude towards it.

When we apply this to a case of failed action, say of someone writing while the ink has run out, then indeed it does not seem to matter whether the thought should fit the world, or the world should fit the thought. Even if the mistake is in the performance, i.e., if the world fails to fit the thought, the thought 'I am writing' fails to be true, and thus the thought cannot be knowledge. Hence Moran is right that an appeal to direction of fit as it is understood in the contemporary philosophy of mind does not help in solving the problem of failed action. However, it is not obvious that Anscombe's claim that 'the mistake is in the performance' should be understood in terms of an appeal to *that* notion of dir-

²⁴⁰He is followed in this assessment by e.g. Grünbaum [2009, pp. 14–15] and Haddock [2010, p. 325]: 'The moral of the story is that if one fails to have knowledge because one's belief is false, the way to ensure that one has practical knowledge is to modify one's behaviour accordingly; there is no need to modify one's belief.'

²⁴¹Also see Velleman [1992a].

ection of fit.²⁴² What if the identification of the requirement of truth for knowledge with the question of what is to be corrected in case of disparity is *not* a confusion? For Moran, there is correspondence or disparity on the one hand, and direction of fit on the other: the first concerns truth or falsity, while the second determines what needs to be corrected, and was therefore mistaken in case of disparity. On an alternative view, being false is already a kind of mistake—one that *only applies* to states and statements with a certain direction of fit.²⁴³

To better understand this suggestion, consider the following. An *assertion* can be mistaken in that it is false, and it can only be called false when it is mistaken. Such is the case with the detective's list: if the detective makes a mistake by writing down 'butter', when in fact the shopper buys margarine, it renders the list false. An *order*, on the other hand, cannot be mistaken by being false. The mistakes one can make while giving orders are of a different sort (inaccurate, inappropriate, inexecutable, etc.). This is the case with the shopping list: the mistake of buying the wrong product is not the kind of mistake that renders the list false—in fact shopping lists aren't the kind of lists that can be false at all.²⁴⁴ Being false thus is a form of being mistaken, but only for certain types of statements (assertions) and states (beliefs), i.e., those statements and states that are supposed to fit the world.

Now what, on this alternative understanding of direction of fit, would it mean to say that 'the mistake is in the performance', and not in the the practical thought? It cannot mean that practical

²⁴²As remarked above, the locution 'direction of fit' does not in fact figure in *Intention* at all.

²⁴³In a later paper, Moran & Stone [2011, pp. 67–69] note that there is indeed a discrepancy between the contemporary notion of direction of fit and Anscombe's doctrine that the mistake is in the performance. However, they do not explain the possible consequences of this fact for our understanding of practical knowledge.

²⁴⁴Imagine the man coming home with the wrong products telling his angry wife: 'Yes I made a mistake, but *your list* is still false!'

thought is like an order and a shopping list in that it is not the sort of thing to which truth or falsity are even applicable, for that would establish that an agent's practical thought is never false only at the price of establishing that it is never true either. And we would again arrive at a funny sort of knowledge. As Anscombe says:

[I]s there not a point at which the parallelism [between an expressions of intention and an orders] ceases: namely, just where we begin to speak of knowledge? For we say that the agent's description is a piece of knowledge, but an order is not a piece of knowledge. So though the parallelism is interesting and illuminates the periphery of the problem, it fails at its centre and leaves that in the darkness that we have found ourselves in. [Anscombe 1957, §31]

Indeed, the difficulty of understanding an Anscombean approach to action lies in conceiving of a kind of thought that cannot be mistaken by being false, but is still potentially knowledge. This then, is the real issue that the problem of failed action brings out. In the next section I will argue that we can come to understand how we can have practical knowledge even when we fail to execute our intention, once we see that our practical knowledge is *constitutive* of what we are doing.

5.6 Practical knowledge: a constitutive account

The previous section concluded that when an agent fails to properly execute her intention, 'the mistake is in the performance' in the following sense: her practical thought is like an *order*, in that falsity is not the kind of mistake that applies to it. On the other hand, we have also seen that practical thought must still be true in order to qualify as knowledge. In this sense, then, it must be

like *belief*. But these two demands seem incompatible. Practical knowledge cannot be like an order because orders are never true. Nor can it be like belief, because belief should represent the world, and is therefore precisely the sort of thing that can be mistaken by being false: the facts are the norm by which belief is judged.²⁴⁵ How are we to unify these two demands?

Consider Anscombe's initial introduction of the concept of specifically *practical* knowledge in *Intention*:

Can it be that there is something that modern philosophy has blankly misunderstood: namely what ancient and medieval philosophers meant by *practical knowledge*? Certainly in modern philosophy we have an incorrigibly contemplative conception of knowledge. Knowledge must be something that is judged as such by being in accordance with the facts. The facts, reality, are prior, and dictate what is to be said, if it is knowledge. And this is the explanation of the darkness in which we found ourselves. For if there are two knowledges—one by observation, the other in intention—then it looks as if there must be two objects of knowledge; but if one says the objects are the same, one looks hopelessly for a different *mode of contemplative knowledge* in acting, as if there were a very queer and special sort of seeing eye in the middle of acting. [Anscombe 1957, §32]

The suggestion is that practical knowledge is different from speculative knowledge in that its character as knowledge is *not* dependent on being in accordance with facts that are 'prior' to the knowledge. This practical mode of knowledge does not fit the propositional attitude model of contemporary philosophy of mind: practical knowledge does not consist of a belief-attitude taken towards a proposition (i.e., a representation that must fit the world by being true).²⁴⁶ Rather, it is the agent's practical thought that

²⁴⁵This might explain the attraction of seeing practical thought as something with a double (pushmi-pullyu) direction of fit, see Millikan [1995, p. 186].

²⁴⁶This fits Anscombe's understanding of practical knowledge as *self-knowledge*, which she notoriously holds to be non-propositional. See Anscombe

she is doing φ that *constitutes* the fact that she *is* doing ϕ .²⁴⁷ Indeed the priority of self-knowledge was already apparent above (§5.2) when we considered that practical knowledge is not derived from the object known but rather is the ‘cause of what it understands’. Consider Elliot again. The question whether she is cooking risotto or spaghetti could not be answered by just observing her movements—we could only see that she was cutting onions. But there *is* a fact of the matter: after Elliot is finished cooking, resulting in a delicious risotto, we can point back to the time when she was cutting onions and say that she was *already* making risotto back then. And Elliot herself was aware of this the entire time. The Anscombean suggestion is that Elliot’s awareness *determines* the fact of the matter that her onion cutting was risotto cooking. What is going on in the kitchen all along falls under the description ‘making risotto’ *because of* Elliot’s thought about what she is doing.²⁴⁸ And if practical thought determines the facts in this way, then of course it cannot fail to agree with them, and hence cannot fall short of being knowledge.

This then is the answer to the question how practical knowledge can be true, without its being able to be false. Elliot’s thought ‘I’m cooking a risotto’ is not true because it correctly *represents* the facts (like a true belief would do), but it is true because it makes it the case that she *is* cooking risotto. It is true in virtue of *constituting* the facts. This explains how the Anscombean fully rules out accidentality. On a causal action theory the intentionality of a movement is determined by something (a mental state, event, or set of mental states and/or events) independent of the bodily movement, and therefore there always is the possibility that mind and movement come apart, as is evidenced by cases of causal

[1975].

²⁴⁷Rödl [e.g. 2007, 2014] further develops the connections between the ideas that practical knowledge is (non-propositional) self-knowledge and that it is a form of knowledge that is constitutive of the facts.

²⁴⁸In this way practical knowledge determines the *form* of what is happening.

deviance. But such accidentality is impossible on the Anscombean view, because the practical thought that determines intentionality is not independent of the movement: it constitutes it.²⁴⁹

But how can this account be correct? Cases of failed action purportedly show that there *can be* disagreement between what the agent thinks she is doing and the facts. Hence in such cases our thought does *not* determine what happens. Or does it?

Suppose that at the time when Elliot is chopping the onions, she is fatally struck by lightning. Then there will never be a time at which Elliot *has cooked* a risotto. But it is nevertheless still true that Elliot was cooking a risotto when lightning struck. As Anscombe says: ‘a man can *be doing* something which he nevertheless does not *do*’ [Anscombe 1957, §23]. This distinction between the *progressive aspect* (doing) and the *perfective aspect* (having done) is widely acknowledged in linguistics, as is witnessed by the large amount of literature on the ‘imperfective paradox’.²⁵⁰ Recently, Michael Thompson [2008] has pointed out the importance of aspect for Anscombe’s account of action. He argues that she conceives of actions as ongoing *processes*, opposed to Davidson’s

²⁴⁹That practical knowledge constitutes intentional action is especially clear when we consider actions that can *only* be performed intentionally, like making a promise or getting married. Anscombe [1963] argues that all the truth-conditions for, e.g., the action-description ‘she is getting married’ may hold except for the fact that the agent does not know that she is marrying (she thinks she is going through a rehearsal). Now what is it that the agent knows in case she *is* getting married? It cannot just be knowledge of all the other conditions for getting married (that witnesses are present, that she says “I do”, etc.) since all these conditions can be in place without her getting married. Rather her knowledge must simply be knowledge of the fact that *she is getting married*. Therefore it is part of the action-concept ‘getting married’, which she applies to herself, that she *knows* she is doing that.

²⁵⁰The imperfective paradox is the problem that we *can* conclude ‘Sam has walked’ from ‘Sam was walking’, but *cannot* conclude that ‘Sam walked to the store’ from ‘Sam was walking to the store’. The latter of course cannot be concluded because Sam’s arriving at the store may have been prevented, e.g., by a lightning bolt. The former, however, can be concluded because whenever there was a period in which Sam was walking, no matter how short, there always will have been a stretch that she has walked. See Dowty [1977].

typical focus on completed *events*.²⁵¹ In Anscombe's examples 'Why?' questions are put to agents in the middle of an action, e.g., 'Why are you replenishing the water supply?' Such statements are not falsified "if philosophers perchance arrange that the H-bomb goes off just now" [Thompson 2011, p. 205] and the water supply is therefore never replenished.

Thompson uses this distinction between the progressive and perfective aspect to answer a challenge related to the problem of failed action: Donald Davidson's challenge that intention does not involve *any* cognitive attitude, be it belief or knowledge, about what you are doing.

[I]n writing heavily on this page I may be intending to produce ten legible carbon copies. I do not know, or believe with any confidence, that I am succeeding. But if I am producing ten legible carbon copies, I am certainly doing it intentionally. [Davidson 1978, p. 129]

Davidson seems to think that an agent's action of making ten carbon copies is necessarily over when she ceases to press down on the paper, at which point she cannot know how many copies have been produced. But Thompson notes that intentionally making ten carbon copies is ordinarily very different:

you write on the top sheet, trying to make a good impression to get through all the carbon, then look to see if your impression made it through all of them. If it did, you stop. If it didn't, you remove the last properly impressed sheet and begin again. If necessary, you repeat. Even the man who has to go through five stages is all along, from the first feeble impression, making ten copies of the document, and he knows it, all along. [...] The one who doesn't know it, Davidson's man, must be under some strange Mafia threat: he gets one chance, no checking, and he's dead if he doesn't manage it. [Thompson 2011, p. 210]²⁵²

²⁵¹Both Jennifer Hornsby [2012] and Helen Steward [2012b] have recently also argued for the need to understand actions as processes rather than events.

²⁵²"There is thus something that Davidson's doctrine of events or of things

Setiya [2012, p. 301] interprets Thompson’s argument as if the point were the following: an agent may believe, or know, what she is doing, as long as she has the opportunity to check her work. The unthreatened office clerk may know what he is doing, and is therefore doing it intentionally, while Davidson’s man is, even if he succeeds, not making ten copies intentionally, because it could only be a matter of luck if ten copies are produced. Setiya correctly argues that this is absurd: surely the second man intends to produce ten copies, and acts intentionally. But Setiya’s reading of Thompson is mistaken. For, as Thompson says, the man who checks and repeats his work knew that he was making ten copies *all along*, from the very start. So even *before* he checked, it was true that he was making ten copies. And this would not have been any different if he had been struck by lightning just as he was about to check his progress: if it was true at t that one was doing φ at that moment, then it cannot at a later moment be false that one was doing φ at t .²⁵³ So the appeal to the progressive should not be understood as claiming that one’s *actual* checking and trying again are necessary for having practical knowledge.

Although some intentional actions may be nearly instantaneous, requiring no checking or retrials, the concept of intentional action would not find application in a world where the contrast between being-in-progress and being complete simply did not exist. That is why Thompson says that ‘Davidson’s man’—that is, a man in a putative Davidsonian world, devoid of progressive goings-on—is not acting intentionally. Intentional action is essentially a matter of doing one thing *by* doing something else.²⁵⁴ It

that happened is missing, namely, not to put too fine a point on it, the things that didn’t happen. That is, he forgets about the things that didn’t happen, but were happening” [Thompson 2011, p. 205].

²⁵³Claiming that Elliot was never cooking risotto, if no risotto gets actually produced, is simply failing to take the progressive seriously.

²⁵⁴As we have seen above (§5.1) answers to the ‘Why?’ question have the form “I am *A*-ing because I am *B*-ing”.

must therefore admit of an internal structure that instantaneous happenings cannot possess.

Once we see that actions are essentially ongoing happenings, the problem of failed action vanishes. Seeing that someone is moving an empty pen over a piece of paper does not allow us to conclude that she is not intentionally writing, for the content of her practical knowledge is “something present, and thus something of which more is to come, perhaps including several attempts at it” [Thompson 2011, p. 210]. In fact we have all had the experience of trying to jot something down with an empty ballpoint, fluently continuing by discarding it for a proper one. Therefore an agent may know what she is doing, even if her *present* attempt to attain her end is failing. This response to the problem posed by failed action thus *denies* that in the cases we have discussed, the agent is ‘flatly wrong’ about what she is doing. Therefore, Anscombe’s positive characterisation of the question ‘Why?’ as pertaining to happenings of which the agent has practical knowledge is saved. In case an action fails, we do not require any ‘weaker description’ of what the agent is doing to stop the means-ends chain from collapsing, for the description ‘I am writing’ is *not* falsified by the fact that the pen has run out of ink.

But doesn’t this outright denial that agents can be flatly wrong about what they are doing lead to contradictions? Suppose an agent, on her way to dinner, knows that the restaurant she wants to go to is located to the north of her current location. Unfortunately, her compass malfunctions, and the agent is led to believe that north is south and vice-versa. Confident that the restaurant is *that way* (the agent points to the direction indicated by her compass), she turns about and starts marching. Now, on the constitutive account, the agent knows that she is going north—that is what she is up to. So the event of the agent’s walking is truly described as ‘walking north’. But it can clearly also be objectively

described as ‘walking south’. Hence, it seems someone can at the same time be doing both p and not- p .

It is indeed true both that the agent is walking north and that she is not. However, this fact does not constitute a contradiction. There is an overarching process of the agent’s ‘going north’ (or more accurately: ‘going north to go to the restaurant’) of which what is happening now—the agent’s going south—is merely a stage. Compare a case in which the agent, to reach her destination, must take a slight detour: imagine that she must cross a bridge down south to make her way north. The agent could say, without the threat of contradicting herself: ‘First I had to turn this way, then that way again – but all the time I was going north, where the restaurant is.’ In this case, walking south is the agent’s *means* to walking north: she knows that by first going south, she will eventually be able to reach her northern destination. It would be absurd to deny that the agent is going north while walking towards the bridge—as if one cannot be travelling to Paris all day, because the road occasionally bends towards Berlin. It may be objected that things appear different in the original case. There, the agent’s walking south is seemingly not a means for reaching the restaurant. If we were to ask her ‘Why are you going south?’, she would at first deny that she is doing that. Indeed, under the description ‘walking south’, what is now happening is *not* something the agent is intentionally doing. But under the description ‘going *this way*’ (where the agent ostensibly identifies what is actually south), it most certainly is. (‘Why are you going this way?’ ‘Why, to reach the restaurant of course!’) It is just that, as things turn out, walking *this way* is not a very good means of reaching the restaurant. This kind of error occurs so often that it is more the rule than the exception. While acting, we constantly correct our movements as we find out that they are currently not sufficient for obtaining our goal—exactly like when we fluently dis-

card an empty pen for a new one to continue writing. Hence, we can dismiss the concern that the constitutive account of practical knowledge would lead to contradictions.

Another worry about restricting the scope of practical knowledge to the progressive, is that practical knowledge would no longer be knowledge of what is actually happening. For instance, Schwenkler [2011, forthcoming] notes that appealing to the progressive dangerously inclines one to think that intentional actions need not involve ‘worldly happenings’ at all. He argues that to say that an agent can know that she is writing even when no such words are appearing as of yet, is to say that she does not need any knowledge of what is actually happening right now in order to know what she is doing. Hence Schwenkler worries that the constitutive account of practical knowledge is committed to a special kind of purely mental “doing in the intentional sense.” But as soon as we accept this kind of dualist activity we can no longer claim “that human actions are events every bit as “worldly” as the rolling of a stone and the fall of a sparrow” [Schwenkler 2011, p. 137]. Indeed it seems that if we were to adhere to this special kind of doing, we would reintroduce a gap between world and thought, and hence could no longer claim that we *do* what *happens*”. Anscombe herself shortly discusses and rejects the possibility of this special kind of doing:

Another false avenue of escape is to say that I really ‘do’ in the intentional sense whatever I think that I am doing. E.g. if I think I am moving my toe, but it is not actually moving, then I am ‘moving my toe’ in a certain sense, and as for what *happens*, of course I haven’t any control over that except in an accidental sense. [Anscombe 1957, §29]

However, it is important to note that this ‘doing in the intentional sense’ which Anscombe rightly rejects is very different from the sense in which someone can truly be said to be doing something in the progressive. This is especially clear if we consider that the

openness of the progressive is not just a feature of intentional actions. As Anscombe says, that an agent can be doing something even though she never gets it done “is in no way peculiar to intentional action; for we can say that something was falling over but did not fall (because something stopped it)” [Anscombe 1957, §23]. Of course the fact that something was falling, without ever having fallen, obviously does not imply that it fell in some mysterious dualist sense. Furthermore, when a sparrow *has* fallen (or an agent has successfully φ 'd), the relation between this event and the earlier *falling* of the sparrow (or the agent's earlier activity of φ 'ing) is obviously *not* accidental. Hence the constitutive account of practical knowledge is in no way committed to the introduction of a dualistic kind of doing. Rather, the omnipresence of progressive judgements ('that stone *is rolling*', 'this sparrow *is falling*') should urge us to take seriously the thought that ongoing *processes* are part of the furniture of the world.²⁵⁵

A final worry about the constitutive account of practical knowledge is that if an agent's practical thought determines what happens, then it seems that agents can be doing whatever they think they are doing. But of course, it is not the case that one can simply be doing anything 'at will'. One cannot count as flying to the moon just in virtue of having an idle wish to do so. Recall that practical knowledge is not just knowledge that one is doing something, but knowledge of an entire structure of means and ends contained within the action. Hence the agent must at least know something about how she can achieve her ends for the action to get started, and thus for practical knowledge to be acquired. You can only know that you are cooking a risotto if you know where to start—e.g., if you know the recipe, or at least know where you can find a cookbook or website that contains the recipe.²⁵⁶ Hence

²⁵⁵Much more on this will follow in Chapter 6.

²⁵⁶Furthermore one might need observational knowledge in cases the description under which an agent acts intentionally includes demonstrative reference

we cannot acquire practical knowledge that we are flying to the moon, simply because we would not know where to start.²⁵⁷ It is essential to the constitutive account that practical knowledge is a *productive* mode of thought, that is, a capacity for knowledge that is only exercised in the production of intentional movements. Indeed, if we are not to fall into the trap of positing a ‘doing in the intentional sense’, acquiring practical knowledge and acting intentionally must be inseparable. We *do* what *happens*.

5.7 Concluding remarks

We would lack free will in a deterministic world, incompatibilists think, because there we would not determine our own behaviour—we would lack self-determination. But the denial of determinism does not ensure, by itself, that we determine anything. So an incompatibilist who affirms freedom—conventionally called a libertarian—must say what more is needed besides the absence of causal determination to get “self-determination.” [Watson 1982, p. 9]

I believe that Anscombe’s account of practical knowledge provides the account of self-determination the libertarian needs. We determine what happens by acting. But this determination is not just the production of some merely physical happenings. It rather is the production of processes that have a rational *form*, that is, a structure of means and ends. And it is by understanding this structure of reasons contained within our actions that we produce them. Hence, Timothy O’Connor is not wrong when he claims

to an object, e.g., ‘to eat *this* apple’. Thor Grünbaum [2012] argues that this means that practical knowledge itself must be observational in some cases. But we can now see that this capacity to make demonstrative reference to an object is, like the relevant knowledge-how, a prerequisite of acquiring an intention with demonstrative content, instead of a source of an agent’s practical knowledge. An agent must use her senses to identify the apple she wants to eat, but once she decides to eat it, her practical knowledge that she is doing so is non-observational. Cf. [Anscombe 1957, §29].

²⁵⁷We will come back to this point in the next chapter (§6.6).

that reasons structure our power to act.²⁵⁸ However, reasons do not structure it by merely determining the probability that we will exercise this power. The power to act rather is a power to produce a rationally structured happening by recognising this rational structure.

By no means is this an easy thought to grasp. And one can justifiably ask how there can be rationally structured happenings and how there can be powers that produce them. Recall John Bishop who warns libertarians about undermining causal action theory, for they would be “undermining the very theory that offers the best prospects for defending the natural possibility of action at all!” [Bishop 1989, p. 66]. Thus one challenge remains for the libertarian that endorses that practical knowledge is constitutive of action: she has to explain how that account is metaphysically possible. This is the final obstacle that stands between us and a proper understanding of the simple picture of free will. In the next chapter I attempt to clear the way by defending an account of powers that can accommodate for the capacity of practical knowledge.

But, given the context of the contemporary free will debate, another somewhat less fundamental question remains as well: is the constitutive account of practical knowledge compatible with determinism? We will be able to say a bit more on this question in the next chapter once the metaphysics is in place (§6.7). For now, though, we can say the following: unlike causal action theory, Anscombe’s account is not *inherently* compatibilist. The control provided by CTA is merely causal control, and causation can be both deterministic and indeterministic. Whether the self-determination Anscombe describes is compatible with determinism is doubtful. Consider the nature of practical knowledge. As we have seen, practical knowledge is not derived from the objects

²⁵⁸See §4.2.3.

known. It rather is a *spontaneous* kind of knowledge: its object, the action, only comes to exist when the agent, by knowing it, determines it to exist. If we take this kind of determination seriously, then we need to recognise that the rational structure we are conscious of in practical knowledge is not merely something that we can project on a series of physical happenings whose existence was fated since the beginning of time. Otherwise human beings would not be able to truly do anything, for they would be no different from Wittgenstein's windblown leaves.

Chapter 6

Powers, Processes and Practical Knowledge*

Dispositions are as shameful in many eyes as pregnant spinsters used to be—ideally to be explained away, or entitled by a shotgun wedding to take the name of some decently real categorical property. It is time to remove this lingering Victorian prejudice. Dispositions, like unmarried mothers, can manage on their own.

—D.H. Mellor²⁵⁹

On the simple picture, free will is the agent's capacity to actualise one of the real possibilities open to her. As we have seen, that capacity cannot merely consist in the agent, or her mental states, *causing* the actualisation of one alternative. For that could happen merely accidentally, and then which alternative gets realised is not up to the agent; it is up to luck. Hence, free will needs to be a capacity to *control* which possibility gets realised. The majority of

*Parts of this chapter were developed in co-operation with Florian Fischer.
²⁵⁹Mellor [1974, p.157].

this thesis has been spent on trying to home in on a suitable notion of control. I have argued that the control required for free action is nothing other than the control exhibited in intentional agency (2). In Chapters 3 and 4 I have considered the most prominent contemporary theories about (free) action and argued that none of them is able to satisfactorily rule out accidentality. The one theory that is able to do this, I suggested in the previous chapter, is Anscombe's account of intentional action in terms of practical knowledge. Because practical knowledge is constitutive of intentional agency, the capacity for practical knowledge *is* the capacity to act intentionally. Furthermore, since practical knowledge is knowledge of the rational means-ends form of action, the capacity to act is a capacity to produce rationally structured processes. This chapter will investigate how there can be such a capacity and how there can be (rationally) structured processes.

To understand the capacity to act we need to understand what capacities in general are—or to use Locke's term, we need to understand what *powers* are. Therefore this chapter will start out with a discussion of the contemporary debate on the metaphysics of powers. Philosophers have found it notoriously difficult to account for the existence of powers. In section 6.1 I will explain the difficulty and hypothesise that it persists because analytic philosophers are too influenced by Humean dogmas. There are two main rival accounts in the debate on powers: reductionism and realism. In the next two sections (§6.2 and §6.3), I will argue that, despite their differences, both positions succumb to the same problem: they fail to account for certain cases where the manifestation of a power is prevented. This problem is widely recognised in the literature on powers, but so far no satisfactory solution has been reached. In section 6.4 I will argue that the problem of prevention has not yet been overcome because both parties in the debate conceive of power manifestations as *Humean events*. If we instead

think of manifestations as *processes* the prevention problem can be overcome. Section 6.5 will explain in some detail what processes are and show how the Anscombean account of agency nicely fits the ontology of processes. Next I will explain how we should think of indeterminism if we accept the metaphysics of powers and processes put forward in this chapter and explain how the simple picture of free will naturally follows from this metaphysics (§6.6). Finally I will discuss whether the combination of an Anscombean account of agency and a metaphysics of processes leaves any room for compatibilism (§6.7).

6.1 Contemporary metaphysics of powers

Ascriptions of powers abound in both everyday life and science. We say: ‘this wineglass is fragile’, ‘haemoglobin can bind oxygen’, ‘aspirin relieves headaches’, ‘the pebble can roll’, ‘the hippopotamus can hold its breath for up to five minutes’, ‘radium can decay’, ‘magnets can attract paperclips’ ‘she can speak Spanish’, etc.²⁶⁰ By ascribing a power to an object, substance or person, we point out one of its properties. Fragility, the power to break when struck, for instance, is a perfectly normal property, common to wineglasses, earthenware plates and porcelain vases. Despite the ease with which we ascribe and single out these properties, analytic philosophers have found it tremendously difficult to understand powers and to accept their existence. This, I believe, is because of the overwhelming influence David Hume has had, and still has, on analytic philosophy in general and the understanding of powers in particular.

Two Humean philosophical tendencies have specifically influenced the understanding of powers. The first is *empiricism*. Powers are allegedly not as easily observable as so-called ‘categorical prop-

²⁶⁰For an extensive argument that power ascription is indispensable to science, see: Nancy Cartwright and John Pemberton [2013].

erties' —like the wineglass's transparency, or its being 21 cm tall. The fragility of a glass only manifests itself when the glass shatters because it is dropped or struck, whereas, say, the wineglass's transparency is manifest for as long as the wineglass possesses that property. Hume writes:

The scenes of the universe are continually shifting, and one object follows another in an uninterrupted succession; but the power or force, which actuates the whole machine, is entirely concealed from us, and never discovers itself in any of the sensible qualities of body. [Hume 1748/2011, p. 103]

Arguably, the first analytic philosophers that discussed powers were early twentieth century logical empiricists. For them the fact that powers are “concealed from us”, is the main reason to attempt to reduce or explain away power ascriptions.²⁶¹ Logical empiricists believe that a sentence only has meaning if it can be verified observationally and therefore, if power predicates can be employed at all in a meaningful scientific language, they have to be analysed in purely observational terms.²⁶²

The second, and related Humean tendency in analytic philosophy is a general hostility towards *modal connections* in nature. Hume famously remarked that “all events seem entirely loose and separate” [ibid, p. 111]. While Hume might have only drawn an epistemological conclusion from this, i.e., that we cannot find out about modal connections between events, neo-Humean metaphysicians argue for an ontological claim: no event can necessitate another event, nor can any event make it possible that another event

²⁶¹To be honest, I do not see at all why powers would be concealed from us. We often *can* see that an object has a certain capacity, i.e., when we see the glass's delicate design we recognise its fragility and handle it with extra care. And even if we cannot directly see that an object has a certain power we can find out about the power by handling the object and playing around with it. This is what scientists sometimes do when they perform experiments.

²⁶²See Rudolph Carnap [1936] for what is considered as the seminal logical empiricist treatment of dispositions.

occurs. As David Lewis, perhaps the most influential neo-Humean of the twentieth century, states:

[A]ll there is to the world is a vast mosaic of local matters of particular fact, just one little thing and then another. [Lewis 1986b, p. ix]

If local matters of fact are the only things that really exist, then all other facts must somehow supervene on them. Lewis calls this ‘Humean Supervenience’, “in honor of the the greater denier of necessary connections” [ibid, p. ix]. The task of the neo-Humean metaphysician is then to explain how all other facts are grounded in these basic local matters of fact.²⁶³ Since powers are properties that specify what is possible (the hippo *can* hold its breath for 5 minutes) or necessary (the glass *must* break if it is subject to a certain force), powers are among the things that supervene on the local matters of fact. Hence if neo-Humean metaphysics can account for powers at all, they must be analysed in terms of the basic local matters of fact.

The tendency to be suspicious of both the unobservable and the modal leads many analytic philosophers to the belief that an object’s capacities are somehow less real than its categorical properties. Especially since powers can remain fully hidden and never actualise: a fragile glass can be manufactured, but before it ever breaks, it might be molten down again.²⁶⁴ Hence most of the work on powers in analytic philosophy is focused on reductively analysing them. However, as we will see shortly, these reductive analyses have been far from satisfactory (§6.2). Moreover, in the last two

²⁶³Perhaps the most famous exemplification of Lewis’s neo-Humean project is his attempt to explain modality itself in terms of local matters of fact. He believes that modal facts about entities in our world are grounded in facts about counterparts of these entities in other really existing possible worlds. Ironically, this reduction of modality to particular matters of fact at other places in the multiverse is known as ‘Modal Realism’. See [Lewis 1986a].

²⁶⁴As Nelson Goodman remarks, capacities strike analytic philosophers “by comparison as rather ethereal” [Goodman 1955, p. 40].

decades more and more philosophers have shifted towards anti-Humeanism. These philosophers explicitly endorse modal connections between events in nature, and accept that powers are as real as categorical properties. Nevertheless, we will see that these new realists about powers struggle with many of the same problems as the reductionists (§6.3). The reason for this, I will argue, is that they still have not overcome all of their Victorian prejudices (§6.4).

6.2 Reduction and prevention

The powers that have received the most scrutiny in twentieth century analytic philosophy are so-called *dispositions*: powers that need to be triggered by a stimulus in order to be exercised.²⁶⁵ Although free will is not a disposition because it can manifest spontaneously, it is still helpful to consider the philosophical treatment of dispositions in order to understand the nature of powers in general. Neo-Humeans have attempted to reductively analyse dispositions in terms of *counterfactual conditionals*. When we say that the wineglass is fragile we are not ascribing a real property to it, we rather just mean that it would break if it *were* struck, even if it is never actually struck. Or more generally:

An object has the disposition to *M* when *C*, if and only if it would *M* if it were the case that *C*. Where *M* is a particular manifestation, and *C* a specific stimulus condition.²⁶⁶

This so-called simple conditional analysis can of course only be satisfactory for the neo-Humean if it is complemented by an account

²⁶⁵Here ‘disposition’ is used as a technical term. In ordinary language we can say things like “the baby is disposed to grow”, or “the teacher has a generous disposition”. To me it seems that neither the baby’s disposition to grow or the teacher’s generosity need a trigger to be exercised.

²⁶⁶Versions of this analysis have been endorsed by, among others, Gilbert Ryle [1949], Nelson Goodman [1955], Wilfrid Sellars [1958] and W.V.O. Quine [1960].

of how counterfactuals are grounded in local matters of fact.²⁶⁷ But we do not need to go into that detail here. For there are counterexamples to the simple conditional analysis.

The first to raise such counterexamples was C.B. Martin [1994]. He exploited the fact that objects can gain or lose dispositions. A copper wire, for instance, can conduct electricity when it is live, i.e. when it is carrying electrical current, but when it becomes dead it loses its disposition to conduct. Now Martin imagines a device he calls an electro-fink, that attaches to a live copper wire and monitors whether a conductor touches that wire. When the electro-fink senses that the wire is touched, it instantaneously stops the flow of electricity to render the wire dead. According to the simple conditional analysis, to say that a wire has the disposition to conduct electricity when touched by a conductor, just is to say that it would conduct electricity if it were touched by a conductor. However, in Martin's imagined case, the copper wire is live and hence has the disposition to conduct electricity, but due to presence of the electro-fink it is false to say that it would conduct electricity if it were touched by a conductor. Thus we have a counterexample to the simple conditional analysis.²⁶⁸

Neo-Humeans have reacted to Martin's counterexample with attempts to modify the simple conditional analysis.²⁶⁹ David Lewis himself, for instance, argues for a more sophisticated conditional analysis. He observes that dispositions have a 'causal

²⁶⁷On Lewis's view, for example, a counterfactual is true just if, in the closest possible world(s) where its antecedent is true, its consequent is also true. Hence the glass is fragile if it breaks if it breaks in the closest possible world(s) where it gets struck. See [Lewis 1973].

²⁶⁸Strictly speaking this is a counterexample to the left-to-right reading of the simple conditional analysis. For the purposes of this chapter, we will not consider the right-to-left reading, and the counterexamples against this. For more on these counterexamples, so-called 'mimickers', see Lewis [1997], Choi [2005] and Fara [2005].

²⁶⁹Lars Gunderson [2002] and Sungho Choi [2006, 2008], have tried to defend the simple analysis as it is. See Manley and Wasserman [2008] for a criticism of these attempts.

basis'. The wineglass's fragility, for example, is based in its micro-structure. Similarly, the wire's disposition to conduct is grounded in its liveness.²⁷⁰ With the help of this observation Lewis proposes to extend the conditional analysis by, among other things, adding a condition concerning the causal basis of the disposition. Informally, this condition states that a manifestation will occur if and only if the disposed object retains the the causal basis of its disposition for some time after it is brought in the stimulus condition. With the help of this condition, Lewis can directly escape Martin's counterexample. For the electro-fink takes away the wire's liveness, i.e. the causal basis for the wire's disposition to conduct, at the moment it is touched by a conductor. Hence Martin's wire does not satisfy Lewis's condition. Although Lewis's sophisticated conditional analysis is able to deal with Martin's counterexamples, Lewis admits that the addition of extra conditions results in quite an "unlovely mouthful" [Lewis 1997, p. 157].²⁷¹

Of course, the complexity of Lewis's analysis would be acceptable if it were free from further counterexamples. But unfortunately for Lewis, this is not the case. Alexander Bird [1998] shows that the sophisticated conditional analysis is unable to deal with *antidotes*.²⁷² Consider arsenic, a poison that is disposed to kill human beings if they ingest it. Even though arsenic has this disposition, someone who ingests a lethal dose of arsenic can still be saved if the right antidote is timely administered. The antidote for arsenic is dimercaprol, which, as Bird remarks, is amusingly

²⁷⁰I will here ignore the question whether all dispositions require causal bases.

²⁷¹To illustrate, here is Lewis's sophisticated conditional analysis in all its complex glory: "Something x is disposed at time t to give response r to stimulus s iff, for some intrinsic property B that x has at t , for some time t' after t , if x were to undergo stimulus s at time t and retain property B until t' , s and x 's having B would jointly be an x -complete cause of x 's giving response r " [Lewis 1997, p. 157].

²⁷²Or *maskers* as Mark Johnston [1992] calls them. While Johnston's paper was published prior to Martin's, its relevance to the debate on dispositions was only recognised after the publication of the latter.

also known as British anti-Lewisite.²⁷³ And indeed the example of arsenic and its antidote works against Lewis's analysis. For dimercaprol does not work by altering any of the intrinsic properties of the ingested arsenic and hence arsenic does retain the causal basis of its disposition to kill.²⁷⁴ Nevertheless, when dimercaprol is administered the manifestation of arsenic's lethal disposition, i.e., the death of the one who ingested it, is prevented. Bird's counterexample illustrates that the manifestation of a disposition can be prevented even if the causal basis of that disposition is retained, and hence this sort of prevention is not ruled out by Lewis's extra condition. Furthermore, Bird's counterexample is a lot less contrived than Martin's, which makes it all the more pressing for proponents of the conditional analysis.

Neo-Humeans have generally tried to rule out Bird's counterexample by proposing even more sophisticated analyses.²⁷⁵ Not only did this result in even unlovelier mouthfuls that are a lot harder to swallow, each new analysis has also been subject to new counterexamples. Hence, this particular episode in analytic philosophy has "endowed us with a delightful bestiary of counterexamples" [Bonevac *et al.* 2012], including: glasses packed in styrofoam, shy but intuitive chameleons, kazoos with wax coated reeds, sorcerers with a knack for protecting fragile glasses, the hater of styrofoam, killer yellow, tricky triangles and concrete blocks with Achilles's heels. As witty as some of the arguments and analyses are, this neo-Humean cottage industry has not resulted in any

²⁷³Lewisite was an arsenic based chemical warfare agent used in World War II.

²⁷⁴Arsenic is poisonous because it binds to certain enzymes and thereby prevents cellular metabolism. Dimercaprol works by binding even stronger to arsenic: it binds yet unbound arsenic and it binds to arsenic molecules that have already bound an enzyme, thus freeing up that enzyme. Subsequently the bound arsenic is excreted in the urine. For a discussion of this process and the question whether arsenic really is an 'antidote' in Bird's sense, see Bird [2007, p. 27,fn. 25].

²⁷⁵See, e.g., Malzkorn [2000], Mellor [2000], Fara [2005] and Manley and Wasserman [2008].

consensus on a satisfactory conditional analysis. In fact, there are nearly as many proposed variants of conditional analyses as there are papers on the subject matter and there does not even seem to be any agreement about the rough direction in which a solution should be searched for.²⁷⁶

6.3 Realism and prevention

The failure of finding a satisfactory reductive analysis of dispositions has prompted some philosophers to propose a radically different account: realism about dispositions.²⁷⁷ This section investigates whether their account of dispositions is more successful than the reductionistic accounts discussed above.

Realists believe that objects have irreducibly dispositional properties that are no less real than categorical properties.²⁷⁸ Therefore, realists, unlike those who attempt a conditional analysis, do not have to ground the existence of a disposition in its manifestation. As Stephen Mumford remarks: “A realist theory of dispositions allows that a disposition’s existence is logically independent of the occurrence of any test or manifestation” [Mumford 2001, p. 377]. But in what, then, does a dispositional property consist? The realist answer is that a disposition is the power to *produce* a manifestation. Fragility, for instance, is the power that *makes* something break when it is struck. But how does that work? According to realists powers bring *de re* connections to the world:

[categorical] properties are intimately connected with powers and

²⁷⁶For an excellent overview of the recent proposals and their problems see [Cross 2012].

²⁷⁷See, e.g., Harré and Madden [1975], Mumford [1998, 2004], Ellis [2001], Molnar [2003], Heil [2005], Martin [2007], Chakravartty [2007], Marmodoro [2010] and Hüttemann [2013].

²⁷⁸Some have even fully turned the table and argue that all properties in fact are powerful properties. For more on such a “pandispositionalism”, see, e.g., Mumford [2004], Chakravartty [2007], Whittle [2008], Marmodoro [2009] and Groff [2013].

have *de re* connections with other [categorical] properties in virtue of those powers. [Mumford 1998, p. 170]

This acceptance of *de re* connections directly opposes the Humean idea that everything is loose and separate. Therefore this realism would not have been possible in the first half of the twentieth century. However since the 1970s it has become more and more acceptable among analytic philosophers to dismiss Humean suspicions about modality. As Stathis Psillos observes:

It was Kripke's liberating views in the early 1970s that changed the scene radically. By defending the case of necessary statements, which are known a posteriori, Kripke [1972] made it possible to think of the existence of necessity in nature [...] [Psillos 2002, p. 161].

Kripke famously argued that although scientific investigation is needed in order to find out that water is H₂O, this nevertheless is a matter of natural necessity—something that is not H₂O cannot be water. The renewed anti-Humean acceptance of necessity in nature, paved the way for contemporary dispositional realists. For they argue that the *de re* connections powers bring to the world, are nothing other than connections of natural necessity:

De re necessity, means necessity in nature: in things, rather than in words or logical form. It is the kind of necessity that is denied by Humeans and by *Tractatus*-Wittgenstein. They think there are no necessary connections between distinct existences. [Mumford 1998, p. 166]

Hence the realist defends that the fragility of a glass is the property that *necessitates* that the glass breaks when it is struck. Brian Ellis, a central proponent of dispositional realism, expresses this idea very clearly. He believes that there are *natural kinds of processes*, and that it is of the *essence* of such processes that one event will lead with necessity to another:

Suppose, for example, that p is a natural dispositional property that would be triggered in circumstances of the kind C to produce an effect of the kind E . Then the processes of this kind will themselves constitute a natural kind, the essence of which is that it is a display of p . Therefore, [...] for all x , necessarily, if x has p , and x is in circumstances of the kind C , then x will display an effect of the kind E . [Ellis 2001, p. 286]

Hence, dispositional realists also accept counterfactuals such as: ‘if an object has the disposition to M when C , then it would M if it were the case that C ’. But they do not think that disposition ascriptions can be reduced to such counterfactuals. Rather, they believe that the truth of such counterfactuals is explained by the existence of powers. An object would M if it were the case that C , because it has a property which necessitates that it will M in conditions C .

Unfortunately for the realists, this reversal of the order of explanation does not allow them to escape all of the reductionist’s problems. If dispositions necessitate their manifestation whenever they are triggered, then how can the realist account for Bird’s antidotes or Martin’s finks? Arsenic’s disposition to kill, for example, does not *necessitate* the death of the one who ingests it—dimercaprol, after all, can prevent lethal consequences. As Markus Schrenk has recently argued:

[I]f [Bird’s] counterexample is indeed effective against the Humean reductionist analysis, then a certain kind of anti-Humean realism about dispositions is also in jeopardy. More specifically, if this counterexample works then it shows *en passant* that metaphysical necessity can hardly be the driving force behind dispositional powers. [Schrenk 2010, p. 729]

Hence, Schrenk concludes that natural necessity cannot play the role the dispositional realist wants it to play. Kripke may have made made necessity in nature philosophically acceptable again, but he only argued for the necessity of *synchronic* identity state-

ments like ‘Water is H₂O’. There cannot be any temporal separation between the property of being water and that of being H₂O—the one cannot come after the other. The dispositional realist, on the other hand, wants necessary connections between temporally separated events. First there is a stimulus then their is a manifestation. But this *diachronic* sort of necessity cannot exist for if two things happen after each other there always is the in principle possibility that something intervenes in between them.²⁷⁹ We can conclude that the counterexamples to conditional analyses of dispositional properties, also form a problem for dispositional realists. Neither the neo-Humean nor the anti-Humean seems to know how they should handle situations where a disposition is in the stimulus condition but its manifestation is prevented.

Recently, Stephen Mumford and Rani Anjum [2011] have argued that the realist can escape the prevention problem by giving up the thought that dispositions necessitate their manifestation. Instead, dispositions give rise to a *sui generis* sort of modality that they argue is “not reducible to pure necessity or pure contingency. It is something in between” [ibid, p. 175]. This *dispositional modality*, the idea is, has some modal strength to produce manifestations in the right conditions, but because manifestations can be prevented it does not necessitate their occurrence—dispositions only display a *tendency* towards their manifestations.

I think Mumford and Anjum’s suggestion is problematic. Not only is it philosophically dubious that there is room in between possibility and necessity for a unique third kind of modality, in many situations it also seems false to say that a disposition only tends towards its manifestations. Consider a wineglass that is crushed under a ten ton concrete block. Do we really want to say

²⁷⁹This problem with diachronic necessity is not new. Anscombe [1971] already argued against a widespread idea that causation is necessitation in a similar fashion. Causation cannot be necessitation, because something can intervene on the causal chain that leads from cause to effect, preventing the effect from coming about.

that this glass only *tends* to break if the block is dropped on top of it? Even if philosophers can imagine a sorcerer that somehow magically saves the glass, then did the glass really remain undamaged because *it* only tended to break? If nothing interferes, then it seems that the glass's fragility *does* guarantee that it breaks under heavy pressure.²⁸⁰ Luckily, we don't need to consider Mumford and Anjum's suggestion any further because giving up the idea that dispositions give rise to necessary connections in nature is, as I will argue in the next section, quite unnecessary.²⁸¹ Prevention cases only present a problem to dispositional realists because they have not fully shed the yoke of Humeanism. Although realists dismiss the dogma that nature is void of modal connections, they still think of the world as a succession of events, or, as Lewis put it, they think of the world as "just one little thing and then another" [Lewis 1986b, p. ix]. Things go awry only when natural necessity

²⁸⁰Mumford and Anjum aim to deal with prevention cases by formalising the strength of tendencies in terms of vectors. A heater has a certain tendency to warm the room, but that effect can be prevented when there is an open window which has a tendency to cool it. Nevertheless, Mumford and Anjum maintain that we cannot conclude anything definitive about what will actually happen by adding up the strength of the different tendencies: "even the resultant vectors can represent nothing more than tendency towards an effect" [Mumford and Anjum 2011].

²⁸¹On the basis of C.B. Martin's mutual manifestation model of dispositions [2007], John Heil [2012] defends another way to deal with prevention cases. On Martin's model, when two dispositional partners meet, they mutually manifest. The wineglasses's breaking is the mutual manifestation of its fragility and, say, a hammer's power to break glasses. Similarly, when a person ingests arsenic the mutual manifestation of the dispositions of arsenic and her body lead to a particular manifestation (the person's death). When dimercaprol prevents this particular manifestation, this is, according to Heil, "simply a matter of the dispositional system that includes the dispositions of your gut, dispositions of the poison, and dispositions of the antidote yielding a different sort of mutual manifestation" [ibid, p. 129]. Hence, it is still necessary for a disposition that it manifests when it is triggered, i.e., meets its manifestation partner, but which manifestation occurs depends on the manifestation partner it combines with. I think that Heil's solution might effectively analyse some cases of disposition manifestation. But it does not work for all problem cases. When the electro-fink kills the live wire, for example, then the wire's disposition to conduct does not yield a different sort of manifestation, it does not yield a manifestation at all.

is supposed to bind these otherwise fully independent happenings together. The time has come to give up this last Victorian prejudice: the picture of the world as a series of separate individual happenings.

6.4 Manifestations as processes

Realists about dispositions deny Hume's dogma that everything is loose and separate. They believe that dispositions are real properties that have the power to produce their manifestation when they are triggered. The production of the manifestation, they believe, can be understood in terms of natural necessitation: when a disposition is brought in the right conditions, it necessitates its manifestation. But the endorsement of natural necessity makes it difficult for the realist to satisfactorily account for cases where the disposition is brought in the right conditions, but nevertheless does not display its manifestation. Dispositions cannot necessitate their manifestations, it seems, because there can always be interference in between two successive events. Markus Schrenk concludes:

[W]e have reasons to be sceptical about the merits of necessitarian dispositionalism, the idea that the link dispositions bring to the world can be identified with metaphysical necessity. [Schrenk 2010, p. 736]

As we saw in the last section, Stephen Mumford and Rani Anjum agree with Schrenk and indeed abandon necessitation in favour of a weaker modal link. But I think that giving up necessitation in the face of the prevention problem is the wrong thing to do. There is another alternative: the realist can give up the thought that the trigger and manifestation are successive *events*. In this section I will flesh out this suggestion and explain how it provides realists with a novel solution to the problem of finks and antidotes.

Many philosophers seem under the impression that everything that happens is just a succession of events, hence they endorse the Humean idea that the world consist of “just one little thing and then another” [Lewis 1986b, p. ix]. However, this event-ontology leads to lots of philosophical problems, one of which, I will argue, is the failure to deal with prevention cases. But what alternative is there to an event-ontology? To answer this question, consider the problem of failed or incomplete actions we discussed in the last chapter (§§5.4–5.6). If the world is just a succession of events, then it seems impossible to account for incomplete actions. Consider Francis who is walking to the store, but never gets there because she is struck by lightning. In this case Francis did not walk to the store, hence there is no *event* of to-the-store-walking. There only was the event of Francis walking to the scene of the meteorological accident. Nevertheless, the imperfective phrase ‘Francis was *walking* to the store’ also has a certain foothold in reality. ‘What was Francis doing when she was fatally struck?’ ‘She was walking to the store.’

According to Michael Thompson Davidson’s theory of action cannot account for incomplete happenings because of its underlying event-ontology:

There is thus something that Davidson’s doctrine of events or of things that happened is missing, namely, not to put too fine a point on it, the things that didn’t happen. That is, he forgets about the things that didn’t happen, but were happening. [Thompson 2011, p. 205]

We could conclude that to correctly conceptualise actions we have to think of them as *processes* that can be ongoing rather than as events that have happened.²⁸² So to answer the question above, the alternative to an event-ontology that conceptualises the world as a succession of things that happen, is a process-ontology that

²⁸²As mentioned before (fn. 251), Jennifer Hornsby [2012] and Helen Steward [2012b] have recently argued for the same conclusion.

accepts the existence of *ongoing happenings*. At this point, the Humean might object that we do not really need to accept processes as a unique kind of entity which is part of fundamental ontology. For Humeans, it seems, can also account for the existence processes by claiming that a process is an event that consist of series of successive smaller events. The smaller events simply are temporal parts of the larger event that is the process. Furthermore, it seems that the Humean can even make sense of the idea that processes can be ongoing, for she can point to an event in the middle of the series and say that the process was ongoing when that event happened. I believe that such a Humean understanding of a process is indeed possible if we consider the process as a completed whole when it has already happened.²⁸³ But the anti-Humean argues that we can also consider a process *while* it is happening:²⁸⁴

our immersion in the flow of time means that we require [...] another way of thinking of occurrents as things which are ongoing, which might turn out one way or another, which might end sooner or rather later, depending on what happens. A process, like a substance, is singled out in thought by way of what might reasonably be called its *form*, which frees it up for participation in such things as growth and change. Just as, in thinking of a horse, say, we lock onto a creature whose principle of individuation has to do with continued life and not with continuation of the very same matter, so in thinking of a process, we lock onto an entity which we conceive of as having a principle of individuation which has to do with what one might call norms of development [Steward 2012b, p. 384]

The *form* of processes will be the topic of the next section (§6.5). For now we will discuss why processes can grow and change, while Humean events cannot.

²⁸³Or when we consider it from what philosophers of time would call an eternalist perspective.

²⁸⁴Or from a presentist perspective.

Humeans typically individuate events based on their spatio-temporal location.²⁸⁵ This means that x is the same event as y if and only if x and y occupy the same region of space-time. Because of this, a Humean event has a fixed duration and cannot grow and become longer, for if it would be longer it would simply be a different event. In contrast, Francis’s currently ongoing walking to the store gets longer and longer the closer she gets to the store until it finally is complete when she arrives there.²⁸⁶ But the process happening right now is the same process as the one that was happening when she opened the door to leave her house, and the same process will still be happening when she enters the mall. Similarly, Francis’s walking to the store can withstand change. Her walking might speed up, for instance, when it starts to drizzle. But the walking process remains the same process as it was before it started raining. A Humean event on the other hand cannot change because it simply consists of the temporal parts of which it is made. As Steward says, a “variation in some quality amongst those parts does not constitute a change in the event itself, any more than a pole’s being red on top and blue on the bottom constitutes a change in the pole” [Steward 2012b, p. 384].²⁸⁷

For the purpose of this chapter, the most important feature of processes is that processes can be cut short before they are completed. Humean events on the other hand cannot be incomplete, for a shorter event would simply be a different event. It was this feature of processes that we used in the last chapter to explain

²⁸⁵Although Davidson openly flirted with the idea of individuating events causally, but ultimately was convinced by Quine to dismiss that idea in favour of spatio-temporal individuation. For this discussion between Quine and Davidson, see Lepore and McLaughlin [1985, pp. 157–162].

²⁸⁶This time there luckily is no thunderstorm.

²⁸⁷In the debate on persistence, this sort of consideration has led to substantive debate. Steward’s argument can be traced back to McTaggart [1927, §§315–316]. See also P. T. Geach [1972, §10.2], Mellor [1981, pp. 110–111], and Simons [1987, pp. 134–137]. It has been attacked by, e.g., Sider [2001, p. 214], who bites the bullet by insisting that “[c]hange *is* analogous to spatial variation”.

how someone can know that she is φ 'ing even if she never successfully φ 's. But as we will currently see, this feature is crucial to understand what happens in cases of manifestation prevention.

The reason why prevention cases are so hard to understand is, I believe, that triggering and manifestations are typically thought of as separate events or “distinct existences” [Mumford 1998, p. 166]. The trigger is thought to be the event that starts a causal chain of events the last of which is the manifestation.²⁸⁸ I would like to suggest that we think of all manifestations as processes, rather than as resultant events. Fragility manifests itself in the *breaking* of the glass and not only in its being broken. We do not only need processes to understand actions—the manifestations of our power to act—we need processes to understand all power manifestations.

As soon as we accept that manifestations are processes we have a natural way of understanding what happens in cases of manifestation prevention. Consider arsenic poisoning again. As soon as a person ingests arsenic, a manifestation process starts. If nothing interferes, the natural end result of that process will be the death of the poisoned person. But the *already ongoing* process can also be stopped short by timely administering dimercaprol. Just as it is true to say that Francis is currently *walking* to the store, even if she never gets there. We say things like: “the arsenic is *killing* her, we must administer dimercaprol quickly!” In fact, we only administer dimercaprol, because we know that a very lethal process is currently ongoing. That process is the manifestation of arsenic's poisonous disposition. Hence we must conclude that antidotes, strictly speaking, do not prevent disposition manifestations at all. An antidote only prevents that a manifestation reaches its end

²⁸⁸Even Brian Ellis, who, as we saw in section 6.2, talks about natural kinds of processes, sees the triggering and the effect as two distinct entities—albeit linked by natural necessity. Cf. “elementary causal relations involve necessary connections between events—namely between triggers and displays of basic dispositional properties” [Ellis 2001, p. 106].

result.²⁸⁹

Once the realist starts thinking of manifestations as processes, it turns out that this case of ‘manifestation prevention’ is not problematic at all. The realist maintains that dispositions, when they are put in the stimulus condition, necessitate their manifestation. The antidote case was supposed to be a counterexample to this thesis, because it was deemed to be a case where a disposition was in the stimulus condition without the occurrence of a manifestation. However, when we give up the idea that the manifestation is the *end-result* of a manifestation process, and instead say that it *is* the manifestation process, we can see that there actually is some manifesting going on before the antidote is administered. And therefore the antidote case in fact nicely conforms to the realist view that the triggering of a disposition necessitates a manifestation-process.

A successful counterexample to the thesis that powers necessitate their manifestations when they are triggered, would be a case where something intervenes *in between* the triggering and the start of the manifestation process. But are such cases even possible? To answer this question, consider the moment at which manifestation processes start. The breaking of a glass starts right when it is struck, the poisoning process starts right when poison is ingested, and a pebble starts to roll right at the moment it is placed on a suitably inclined surface. In other words, a manifestation process starts right at the moment when it is triggered. If this view is correct, then trigger and manifestation are not distinct events. Rather, the triggering of a disposition *is* the start of the

²⁸⁹Hence Heil’s solution to the antidote case discussed in footnote 281 was not fine-grained enough. On his view, a case where both a poison and an antidote are administered yields a different manifestation, than a case where only the poison is administered. But what in fact happens is that we first have a manifestation of the dispositions of the poison and the body—a poisoning process—which is then cut short by the manifestation of dimercaprol’s disposition to bind arsenic.

manifestation process.²⁹⁰

There are two related conclusions to be drawn about the proposed view that manifestations are processes which start with triggers. First, the realist who adopts this view can agree with Schrenk that natural necessity is unfit for binding together temporally separated events. However, this realist can hold that dispositions still bring necessary connections to the world—the fragility of the glass *necessitates* that striking the glass starts a breaking process. But since the moment at which the glass is struck is the moment at which the breaking starts, the necessity involved is *synchronic* rather than *diachronic*. Second, because a manifestation process starts right when it is triggered, there is simply no room for any interference in between the trigger and the start of the process. Therefore finks, antidotes and the like are *conceptually impossible*. Of course this does not mean that there are no devices to kill a live wire, or that dimercaprol does not exist, what I mean is that cases where a disposition is triggered without manifesting cannot occur in nature.

To add some force to this second claim, consider how we can understand Martin’s electro-fink on the proposed view. Martin supposes that the electro-fink “can provide itself with reliable information as to exactly when a wire connected to it is touched by a conductor” [Martin 1994, pp. 2–3]. Furthermore, he supposes that the fink instantaneously reacts when contact occurs. But it is dubious that such a device really can exist, for processing the information that an attached wire is touched must at least take some time. As Lewis observes, to avoid this misgiving we can easily remove the supposition that the electro-fink reacts instant-

²⁹⁰Of course, a disposition is often triggered as a result of a temporally extended triggering process. I am not claiming that a disposition starts to manifest as soon as such a triggering process *starts*. The striking of a glass with a hammer, of course, begins before the glass starts to break. I only want to suggest that the moment at which this striking process triggers the glass’s disposition to break, is the moment at which the glass’s process of breaking starts.

aneously: “Quickly is good enough. Then the electro-fink [...] need not be anything more remarkable than a (sensitive and fast-acting) circuit-breaker” [Lewis 1997, p. 147]. If the electro-fink indeed works as a circuit breaker then it cannot prevent the manifestation process of the wire’s disposition to conduct. In fact a circuit-breaker can only spring into action when it detects that at least some conducting is ongoing. Hence, the circuit breaker, like dimercaprol, does not really prevent manifestation, it rather intervenes and cuts off an ongoing manifestation. Therefore the case of the circuit breaker does not constitute a counterexample to the view that dispositions necessitate their manifestation-processes when they are triggered. However, the electro-fink might also work differently. Sungho Choi and Michael Fara [2012] write that an electro-fink “is a device which senses when the wire is about to be touched by a conductor”. We could imagine that a camera is monitoring the proximity of the conductor to the wire and right before they touch the fink renders the wire dead. Now this would be a genuine case of manifestation prevention, since the manifestation process does not even start. But of course the only reason why the manifestation process does not start is that the disposition to engage in that process is removed *before* it can be triggered. Therefore, the camera operated electro-fink does not present a counterexample to the view that dispositions necessitate manifestation processes *when* they are triggered.

Since it is conceptually impossible that there is any interference in between trigger and manifestation all of the purported cases of finks and antidotes proposed in the literature must be analysable as either cases in which something intervenes on the already ongoing manifestation, or as cases in which the manifestation process is genuinely prevented, but only because the disposition is never actually triggered.²⁹¹ But it goes beyond the scope of this thesis

²⁹¹In the camera operated fink case the manifestation was prevented by taking away the disposition. But there might also be cases where the triggering

to sort out all the examples. This chapter are trying to understand the nature of powers in order to see if we can make sense of the power to act: the capacity for practical knowledge. In this section we have taken the first step towards understanding how this capacity can have a place in nature. For when we discussed in the last chapter (5.6) that practical knowledge essentially is knowledge of an ongoing process, that actions are ongoing processes might have seemed unique or even mysterious—and metaphysicians might be hesitant to accept the *sui generis* ontological category of processes just in order to account for agency. But now we can see that accepting that powers give rise to processes is not an *ad hoc* move we need to make in order to account for the power to act, we need to accept the category of processes in order to understand the nature of all powers. All powers produce processes.

6.5 The form of processes

The identity of processes, unlike the identity of Humean events, does not depend on spatio-temporal location—a process can become longer, larger, and can withstand certain changes. On what then does the identity of processes depend and what determines which changes a process can and cannot survive? As we saw in the last section, Helen Steward suggests that a process can be identified “by way of what might reasonably be called its *form*” [Steward 2012b, pp. 384]. But by itself this suggestion seems quite obscure, for what is a form and how can a process come to have one? Luckily we have already come across the idea that a process can have a form in the previous chapter: actions have a rational structure. Moreover, the structure an action has is determined by practical knowledge or, in other words practical knowledge is

is prevented from coming about. Striking a bubble wrapped wineglass, for instance, does not produce the amount of force required to trigger the glass’s disposition to break.

the *formal* cause of action.²⁹² Hence the idea that processes have forms is not entirely new to us. The last section argued that actions are not special or unique because they are processes, in this section I will argue that actions are not peculiar either in that they have a form. Nor is practical knowledge mysterious because it is a formal cause. All processes have a form and all powers are the formal causes of their manifestations.²⁹³

Let us start by recuperating the idea that actions are enformed processes. Recall Elliot (§5.2). Her walking to the kitchen, taking ingredients out of the cupboard, cutting onions, turning on the stove, etc., are not just a bunch of random movements she makes, they are all part of one overarching process of risotto cooking. They are part of this process because Elliot thinks of all these movements as *means* to her *end* of risotto cooking. In this way Elliot's practical knowledge *unifies* these movements as phases of one action process. Furthermore, Elliot's practical knowledge does not just determine which movements are part of her risotto cooking, it also determines which movements are no part of it. The telephone call she makes while she is waiting for the water to cook, for instance, is no part of her risotto cooking, because she does not conceive of this phone call as something she does in order to cook risotto.²⁹⁴ Hence practical knowledge determines the form of an action, because it is knowledge of the action's aim that conceives of the different action-phases as means to reach that aim. In other words, practical knowledge determines the *rational* structure because it is knowledge of this structure.²⁹⁵

The form of an action processes is its rational structure. But

²⁹²As remarked in footnote 218, 'formal cause' is a term Aristotle uses for one of his four causes.

²⁹³Sebastian Rödl [2012/2005, Chapter 5] has recently argued for a similar conclusion, if I understand him correctly.

²⁹⁴Unless of course, she is calling her father to sort out the precise details of the family recipe for risotto.

²⁹⁵As explained in §5.6, practical knowledge is constitutive of the action.

what is the form of processes that are not rational? One suggestion, by George Molnar [2003], is that the intentionality we find in actions also exists in the purely physical world. How can this be? Beliefs, desires, intentions, and the like are directed at their objects because they *represent* them,²⁹⁶ but there is no representation in the purely physical world. However, against philosophical orthodoxy, Molnar argues that representation is not essential to intentionality. What is essential, rather, is directedness itself. According to Molnar there exists a non-representational sort of directedness, which he calls *physical intentionality*. This sort of intentionality is exhibited by physical powers:

A power has directionality, in the sense that it must be a power for, or to, some outcome. It is this directedness that provides the *prima facie* distinction between powers (dispositions) and non-powers. [ibid, p. 57]

I believe that Molnar's suggestion goes too far. For it seems wrong to say that fragility, for instance, is aimed at the breaking of the glass. In no way is breaking the *goal* of fragility. Whereas it seems correct to say that cooking risotto can be the goal of an intention and the movements it produces. Therefore we should not accept physical intentionality. As John Heil remarks:

“[W]e have Brentano's thesis that intentionality is the mark of the mental turned on its head: intentionality is a mark of the material!” [Heil 2004, p. 440]²⁹⁷

Nevertheless a non-representational form of directedness—whether it merits the name ‘intentionality’ or not—might be present in biological processes. It could be argued that certain powers of living organisms are aimed at the good of the organism. In other words, there might be natural or biological *teleology*. Consider the power

²⁹⁶This was discussed in §3.4.

²⁹⁷It must be noted that Heil's assessment of Molnar's [2003] book is not as negative as it might seem from this quote. Heil points out that the book “represents an important move forward in philosophy.” [Heil 2004, p. 443].

of a tadpole to develop into a full-grown frog. There clearly is a standard way in which the developmental process from tadpole to frog unfolds: first it grows hind legs, then it grows front legs and finally the tadpole sheds its tail. Like the different phases of an intentional action, these phases of frog development are not random happenings, they are unified by the fact that they are all part of the manifestation of a power that *aims* at becoming an adult frog. Furthermore, this power also determines which happenings are not part of the frog's developmental process. The ripples a tadpole causes by swimming around are no part of the process of becoming an adult frog because these ripples do not have this development as their goal. Hence we might say that the power of a tadpole to develop into a full-grown frog is the formal cause of the frog development process, just like we can say that practical knowledge is the formal cause of action.

Perhaps one might be sceptical about natural teleology, as I am sure many philosophers are. But it should be noted that denying natural teleology comes at a price:

In the description of photosynthesis, for example, we read of one chemical process [...] followed by another, and then another. Having read along a bit with mounting enthusiasm we can ask: "And what happens next?" If we are stuck with chemical and physical categories, the only answer will be: "Well, it depends on whether an H-bomb goes off, or the temperature plummets towards absolute zero, or it all falls into a vat of sulfuric acid" [...] That a certain enzyme will appear and split the latest chemical product into two is just one among many possibilities. Physics and chemistry, adequately developed, can tell you what happens in any of these circumstances—in *any* circumstance—but it seems that they cannot attach any sense to a question "What happens next?" *sans phrase*. [Thompson 2008, p. 41]

These "What happens next?" questions, it seems to me, have genuine answers. At least the onus is on her who is sceptical about

natural teleology to present a compelling argument why these questions do not have answers. Alternatively she could take up the difficult task of showing how these questions can be answered purely in the terms of physics or chemistry.²⁹⁸ However, this is not the place to pursue a defence of natural teleology any further.²⁹⁹ For this thesis does not rest on a commitment to natural teleology. The current purpose is to try and understand the forms of processes and their causes. We have seen that the phases of teleological processes, such as intentional action, and—if you are willing to accept natural aims—photosynthesis, are unified by a power that aims at an endpoint. But not all processes have a natural aim so the question remains how we are to understand the form of non-goal directed processes.

Consider a pebble that rolls down a hill towards a river. The first stretch of rolling, say from the top of the hill to halfway down the hill, and the second bit of rolling, from the halfway point to the river, seem part of the same process. But in this case we cannot say that these parts belong to the same process because the two stretches of rolling are means towards an aim. Although the rolling does have a direction (the pebble rolls *towards* the river), the process is not *directed* in the sense that it has a certain end-point, say, reaching the river, as its aim. Hence we cannot ask “What happens next?” questions about the process of rolling: it is

²⁹⁸The latter can, for instance, not be done on the simple basis of statistics. If we consider a newly hatched tadpole and ask what happens next, the answer, according to the elementary school biology book, is that it will grow hind legs. But in reality *most* tadpoles get eaten before they have the chance to develop further. Nevertheless the biology book is not lying when it presents us with the familiar picture of the frog’s development cycle.

²⁹⁹Jesse Mulder [2014, Chapter 7] provides a compelling defence of natural teleology. He also argues that we can only take the idea that biological kinds determine the structure of processes seriously, if we accept that the processes are not fully determined on the level of physics and chemistry. Hence, he denies the compatibility of determinism with any life-processes, not just intentional actions, or animal movements. While I am certainly sympathetic to that view, a defence of such an incompatibilism is beyond the scope of this thesis.

in no sense more natural for the rolling pebble to be rolling all the way to the bottom, than to be stopped by an obstructing object halfway down the hill. Some inanimate processes might have a natural end point, in that they stop when they reach a state of equilibrium for instance. But it would go too far to say that such processes *aim* at the equilibrium. So how then is the identity of inanimate processes determined? I think that the identity can again be grounded in the power to engage in that process. We can think of the two stretches of rolling as part of one process, because the stretches are part of a *continuous manifestation* of the power to roll. If the pebble were to be stopped halfway down the hill by a block of ice, and only would roll from the half way point to the river after spring has come and the ice block has melted, we would have two rollings instead of one. Furthermore, although a rolling process does not have an aim, it does have a certain modal profile. It could, for instance, have lasted longer if the hill was a bit larger. This too can be grounded in the power to roll: the process could have been a little longer, because the power to roll would have continued to manifest if the hill was longer. Finally, the power to roll also determines which happenings are no part of the rolling process. The nearby waves in the river, for instance, are no part of this process, simply because they are not manifestations of the power to roll. Thus we can conclude that even the powers of inanimate objects are the formal causes of their manifestations.

The idea that the identity of a processes is grounded in powers, I believe, finds further support in the way in which experimental science conducts its research. When scientists try to find out something about the powers of a particular object, or the nature of a specific process, they try to isolate a pure manifestation by shielding it off from as many potentially influential powers as possible—they perform experiments in a vacuum chamber, or a cleanroom, for instance. Hence it seems that the identity of a process, in

its purest form, is indeed determined by the power of which the process is a manifestation.³⁰⁰

The above analysis, I hope, makes it plausible that processes have a structural form which is determined by powers. Before we move on to see how indeterminism fits in the picture of powers and processes, I wish to briefly point out one benefit of the idea that powers are formal causes: it offers a very straightforward answer to the classical objection that powers lack explanatory force, the so-called *virtus dormitiva objection*.³⁰¹ This objection derives its name from Molière's play *Le malade imaginaire*. In the play a candidate doctor is asked why opium makes one sleep. The candidate responds in his best high-sounding Latin that opium makes one sleepy because it has a *virtus dormitiva*. With this answer the chorus is very impressed and they promptly welcome the candidate among their highly learned ranks. The joke, of course, is that claiming that opium has a *virtus dormitiva*, just is a maronic Latin way of saying that opium makes one sleep. Triviality can go no further, it seems. And indeed, when someone enquires why opium produces sleepiness, it does not help at all to simply say that opium has the power to make one drowsy.³⁰² However, if one asks a different question, not about opium but about one of opium's effects,

³⁰⁰As we saw above (fn. 281), we can think of some processes as the mutual manifestation of multiple powers. It might even be argued that all powers are like this. The rolling of the pebble, for instance, might be seen as the mutual manifestation of the pebble's power to roll, the earth's gravitational powers and the terrain's power to provide friction (without friction there would be sliding instead of rolling). For both simplicity and brevity I will not discuss such mutual manifestation processes here, but it seems fine to say of such a process that it is formally caused by all of the powers that combine in its manifestation. For the same reasons I will not consider the powers of artefacts, i.e., inanimate but man-made objects.

³⁰¹The objection is pushed in analytic philosophy by, e.g., David Armstrong [1973] and J.L. Mackie [1977]. For an extensive discussion of the problem, see [Mumford 1998, §6.6-6.9] and [Michon 2007].

³⁰²To answer the question we need to look at the process that happens when humans consume opium and find out what it is about the opium, and what it is about humans, that makes opium have this effect.

a question such as “why is this person sleepy?”, then the answer, ‘because of opium’s dormative virtue’ is not explanatorily useless. For at least it explains that the opium induced the sleepiness instead of, say, the monotone voice of the doctor who administered it.³⁰³ But, I believe, the role powers play in causal explanation is not fully discharged in that they exclude other causes.³⁰⁴

When we walk into the kitchen and see a broken wineglass on the floor we can wonder why it broke. One explanation of the broken glass points to the *efficient cause*³⁰⁵ of the breaking, e.g., ‘because it fell from the counter’. Often we would like to probe further and might find out about an earlier efficient cause: the cat pushed the glass off the counter. However, there also is an entirely different, though related sort of causal enquiry that can never be satisfied by tracing back a chain of efficient causes: we could wonder why the glass would break, instead of bounce, when it is dropped. Philosophers of causation often forget about this second sort of enquiry. All they try to understand is the causal relation between subsequent events—another trace of Humeanism, one might suspect. But to someone who, for some reason, is only familiar with cheap plastic wineglasses, the second line of questioning might be the more pertinent of the two. Questions in this line are answered by pointing to the powers of the object. The wineglass breaks when it is dropped, because it is fragile and not flexible. The powers of an object explain why the object reacts in a certain way to certain stimuli—they determine in which sorts of processes the object can engage. In other words, they determine the *form* of the causal processes the object participates in. This is why it is not trivial to point to powers in causal explanation.

³⁰³This defence of powers against the *virtus dormitiva* objection is offered by Keith Hutchinson [1991].

³⁰⁴Of course it is also possible that the sleepiness is brought about by more than one cause, one can be sleepy because one took some opium after a well earned dinner ensuing a hard day of work.

³⁰⁵Another of Aristotle’s four causes.

Of course, we will not often wonder why a glass that is dropped does not, say, turn into a tadpole, but instead starts breaking. But when we encounter new kinds of things we are unfamiliar with, we precisely ask how they respond to certain stimuli, and in what sorts of processes they can engage. And these questions are answered by finding out about the powers of these things.

This chapter asks how it is possible that the capacity for practical knowledge is a capacity to produce structured action processes. In this section I have attempted to show that every process has a form or structure which is grounded in the power to engage in that process. Therefore, if we accept the anti-Humean metaphysics of powers and process, there is nothing overly mysterious about practical knowledge. In chapter 4 we saw that contemporary agent-causalists employ a similar strategy to demystify their notion of agent-causation. They argue that this form of causation does not only occur when an agent acts, rather substance-causation is a ubiquitous phenomenon. The problem with their argument, however, is that by emphasising the pervasiveness of agent-causation they lose the ability to explain what is special about agential-control. Does the Anscombean account of control run the same risk? I do not think so. Practical knowledge remains unique in that it is a *conscious* power, and that is the reason why practical knowledge provides action with a structure that is *rational*. At first sight this claim might not seem much different from the idea of contemporary agent-causalist such as Lowe or O'Connor that an agent's ability to act is a rational power.³⁰⁶ However, the only way in which they could explain the rationality of the power to act was by holding that the power to act does not directly produce movement. It rather produces something that is still fully mental: a volition, or an action-triggering intention. The Anscombean, on the other hand, can explain why actions

³⁰⁶See §4.2.

themselves are rational: they are rational because they possess an internal means-end structure in virtue of the agent's practical knowledge of that means-end structure.³⁰⁷

Although we can now understand how practical knowledge is possible, there is one step left to take in order to fully understand the metaphysics behind the simple picture of free will. For on that picture, our actions are the actualisations of really possible alternatives open to us. In the next section we will investigate where these possible alternatives come from, and how our actions can be seen as the realisations of such alternatives.

6.6 Powers, indeterminism and control

In a short and astute paper, Arthur Prior [1962] discusses an interesting argument of the 18th century American philosopher and theologian Jonathan Edwards. Edwards argues against the possibility of indeterminist free will on the grounds that uncaused actions or volitions lead to absurdity.

What is self-existent must be from eternity [...]: but as to all things that *begin to be*, they are not self-existent, and therefore must have some foundation of their existence without themselves.

[Edwards 1754/2009, p. 181]

If there is an event, say a volition, that does not have a cause which grounds its existence in what comes before it, Edwards argues, then there might as well be many more events like it popping up all over the place. So far, this is not a bad result for Edward's opponent, for she indeed thinks that there are lots and lots of uncaused events, in particular volitions, that occur all the time. Edwards presents an interesting reply to this thought that is, as far as I know, not often discussed in the contemporary free will

³⁰⁷Nevertheless, I think it is correct to say that the Anscombean account of action presented in this thesis is an agent-causal account, despite this difference from other contemporary agent-causal accounts.

debate. He argues that just like there is no reason to suppose that there is only one uncaused event—if there is one, there might be many—there equally is no reason to think that all uncaused events should be of one sort:

If it were so, that things only of one kind, *viz.* acts of the will, seemed to come to pass of themselves; [...] this very thing would demonstrate that there was some Cause of them, which made such a difference between this Event and others. [Edwards 1754/2009, p. 184]

Hence Edwards's idea is that it is absurd to think, like some of his contemporaries, that only volitions can occur without a cause. If there is one uncaused volition, there is nothing that stops all sorts of other events from randomly occurring. But perhaps the advocate of volitions would object that there is something special about the *nature* of volitions that makes them start to exist without a cause. This retort however, will not do:

I would observe, that the particular nature of existence [...] can lay no foundation for that thing's coming into existence without a Cause; because to suppose this, would be to suppose the particular nature of existence to be a thing prior to existence; and so a thing which makes way for existence [...] The distinguished nature of the effect, which is belonging to the effect, cannot have influence backward, to act before it is. The peculiar nature of that thing called volition, can do nothing, can have no influence, while it is not. And afterwards it is too late for its influence: for then the thing has made sure of existence already, without its help. [ibid, p. 185]

Edwards concludes that it is “repugnant to reason” [ibid, p. 185] to suppose that only acts of will are uncaused.

This argument, I think, can be seen as a sort of luck objection on a grand scale. The standard luck objection, discussed in Chapter 2, challenges the thought that an act can be undetermined, for then it would not be up to the agent, it would merely

be a random happening up to nothing but chance. Edwards's argument goes one step further and purports to show that if there is but one uncaused event, then everything might as well be random. And this argument, as Arthur Prior observes, does not only pose a threat to those who believe in undetermined actions: it presents a challenge for anyone who supposes that there are certain events, like the decaying of a radium atom, that happen without a cause. No serious indeterminist, of course, believes that everything that happens is purely random, she still recognises that certain mechanisms might work deterministically, and even the occurrence of indeterministic happenings is not *purely* random—there are statistical laws concerning those. Hence, Edwards's argument challenges not just libertarianism, it challenges the possibility of any acceptable sort of indeterminism.

But is his argument valid? I think it might be valid against someone who thinks of the world as just one long succession of events—as just one little thing followed by another. For then indeed the world would be a continuous coming and going of *fully independent* things and nothing about the earlier things can tell anything about the coming into existence of later things.³⁰⁸ But if we deny such Humeanism, there is a way out of Edwards's argument. As Prior remarks, talking about the happening of events as 'something coming into existence' is not innocent:

[T]his change of key is not metaphysically illuminating but metaphysically obfuscating. [...] we explain what is meant by a

³⁰⁸And indeed it can be argued that Hume himself was a determinist, see Peter Millican [2011]. Neo-Humeans however, have tried to come up with ways of understanding indeterminism in their framework—perhaps motivated by the fact that current science tells us that the world is fundamentally indeterministic. David Lewis [1979], for example, defines that a world is deterministic, if any world with the same laws of nature is either fully alike that world (throughout all of time) or never alike it, and indeterministic otherwise. To me this definition only seems to provide a derivative notion of indeterminism that has nothing to do with the alternatives we see before us when we act. After all, the full history of any given Lewisian world is settled anyway.

headache's starting to exist by saying that it just means a head's starting to ache, not vice versa. Behind this view is, of course, the assumption that the world consists not of events, such as headaches, but of things, such as heads, which interact and change. And a remoter part of this underlying assumption is that how things behave—that is what events occur—is determined partly by their natures or dispositions and partly by what happens to them. [Prior 1962, pp. 58–59]

In a world of substances, powers and processes, nothing stands in the way of indeterminism. It is perfectly possible that some kinds of things, like a radium atom or a human being, have alternative possibilities, where others entities behave fully deterministically. And of course, the entities that engage in indeterministic happenings do exist prior to the occurrence of these happenings. Therefore they can ground what is possible and what is not. A radium atom, for instance, can emit an alpha particle but it is in the nature of radium that it cannot, say, engage in rope-skipping. Hence the powers of objects do not just determine the form of their manifestation processes, they also determine which processes are really possible:

We cannot and do not need to say that it is in the nature of 'volitions', or of certain volitions, which makes their coming-to-be possible but not necessary, and the nature of other non-existent or not-yet existent occurrences which makes their coming-to-be either necessary or impossible. It rather is that there are certain already existing objects which have certain capacities, and some which lack them, and none which have certain other capacities. [ibid, p. 59]

Therefore, Prior argues, only a metaphysics of substances and powers makes it possible to spell out a 'limited indeterminism'—an indeterminism that does not make everything contingent and random.

The simple picture sketches free will as the capacity of human beings to choose which path they will take at different junctions in the universe's garden of forking paths. We can now see how this garden itself is shaped: which paths there are is determined by our own powers, and the powers of the entities with which we co-inhabit this world. What is really possible at a specific time and place is determined by the powers of the objects located there. It is really possible for me right now to break my coffee cup because the cup is fragile, I have the power to throw it and I am sufficiently close to the cup.³⁰⁹

Since our powers together with the powers of the objects around us determine which paths extend themselves before us, exercising a power is like walking on one of the paths. By exercising our capacity to act in one way, rather than in another, we actualise one of the alternatives that previously was merely possible. At the same time this exercise determines that some alternative paths are closed off and are possible no longer. Hence, acting *consists* in directly controlling what possibilities get actualised. As Ruth Groff states, once we see that acting is nothing but an agent's exercise of her capacity to act, "control comes for free" [Groff n.d., p. 14]. There is no question whether what happens is up to the agent, for what happens simply is a manifestation of the agent's power to act. This does not mean, however, that every power manifestation is an instance of control. For else a radium atom would be controlling its decay.³¹⁰ The power to act is special because it is a power to consciously bring about rationally structured processes.³¹¹ In the

³⁰⁹For a detailed account of how real possibilities are grounded in the powers of objects, see [Rumberg n.d.].

³¹⁰Cf. §4.2.1.

³¹¹Hence, a realist metaphysics of powers is not sufficient for understanding control. It must be combined with a fitting account of acting for a reason. Groff is well aware of this, as mentioned before (fn. 184), she writes about acting for a reason: "Of all of the issues on the free will side of the problematic, this one is perhaps most salient for the dispositional realist. I say this because cognition is clearly a core component of the power to spontaneously and intentionally

last chapter I have presented an account of this power to act, and in this chapter I have tried to show how this power nicely fits in our indeterministic world. If my attempt has been successful, then the second objective set for the libertarian in chapter 2 (p. 85) is completed: we have arrived at an account of agency that explains the possibility of action in our indeterministic universe. I believe that if we accept the Anscombean account of acting for a reason, and ground it in a metaphysics of powerful substances, there is no need to adjust or replace the simple picture of free will. The picture can be vindicated as it stands. Thus I hope we have arrived at a “reasonable libertarianism” [Wiggins 1973].³¹²

6.7 Practical knowledge and determinism?

In the previous section we have seen that Anscombe’s account of agential control in terms of the power of practical knowledge nicely fits together with an indeterministic world view. Furthermore, if accepting Anscombe’s teachings about action involves accepting a world of powers and processes, I do not see what motivation remains to insist on determinism. If you reject the idea that the world consists of one large chain of events, then why would you think that the symphony of humans, quarks, animals and planets would, at every point in time, leave open only one, and exactly one, possibility? But even if there remains no clear motivation to insist on determinism, philosophers might still wonder whether or not it is *conceptually* possible to have agential control in a deterministic world. We saw in Chapter 3 that the usual way

exercise one’s other powers. The concept of acting for a reason can therefore be expected to play an important role in any powers-based account of free will” [Groff n.d., p. 21]. In this thesis I started out by looking for a good account of acting for a reason, and found that it fits well with a metaphysics of powers. But I could also imagine that one starts from a powers metaphysics and then finds that an Anscombean theory of acting for a reason fits it nicely.

³¹²See page 13.

in which compatibilists try to spell out the control required for free will, i.e., in terms of event-causation and representation, is hopeless.³¹³ But could a compatibilist not also try to spell out control in Anscombean terms? In this final section I will argue that she cannot.

In the last chapter we saw that even though practical knowledge is spontaneous, we cannot just start to know that we are doing anything whatsoever.³¹⁴ Acquiring practical knowledge—i.e., forming an intention—for instance presupposes *know how*. In order to cook a risotto, we need to know the recipe, or at least we need to know where to find a recipe.³¹⁵ But the knowledge required for intentional action is not limited to know-how. In order to engage in action we also need knowledge that the world will remain relatively stable. Consider the following example by Georg Henrik von Wright:

On the desk in front of me there is now a loose sheet of paper. I know I can turn it over if I want to. Implicit in this knowledge of my ability is my confidence that the paper in front of me will remain in its present position, unless I turn it. Should I exercise my ability and turn the sheet, I could afterwards say confidently that, had I not interfered, the sheet would have remained in its present position. It happens, of course, that I am mistaken. Perhaps a sudden breeze through the room turns the paper in front of me the very moment when I am about to do it myself. But this is exceptional. If such cases were common, they would weaken my confidence in my ability to perform the action. And if it was the rule that sheets of paper turned over and whirled about quite unpredictably most of the time, there would be no

³¹³And hence, if John Bishop is right that compatibilism necessarily presupposes the acceptance of causal action theory, then any compatibilism is doomed to fail. See: §3.5 and [Bishop 1989].

³¹⁴Rather the spontaneity of practical knowledge resides in the fact that it is the cause of the known object. See §5.2).

³¹⁵As we saw in the last chapter (§5.6), the fact that practical knowledge presupposes such know how, is the reason why we cannot simply start knowing that we are flying to the moon.

such action as turning over a sheet of paper. [von Wright 1974, p. 42]

Two conclusions can be drawn from this example. The first is that action would not be possible in a world of *unlimited indeterminism*. The ability to turn over a piece of paper would not exist at all if papers were always randomly turning over of their own accord. Hence, in a world where everything that happens is contingent and random there would not be any action whatsoever: “Action can thus be said to presuppose regularities in the world” [ibid, p. 42].

The second conclusion to draw is that action would not be possible in a totally deterministic world either. For if acting involves confidence that the world will turn out different if we do not act—that the sheet of paper will not move unless we act—then every action presupposes at least two different possible ways in which the world may develop: one where we act and one where we refrain from acting. As von Wright recognises, we are sometimes mistaken that the paper would remain where it presently is, in case there is a sudden gust of wind, for instance.

But, by and large, the type of confidence to which I am here referring is trustworthy. Were it not so, *action* would not be possible. It is a conceptual feature of the utmost importance of that which we call action that certain changes in nature would not have occurred had we not produced them—or would have happened had we not prevented them. [von Wright 1973, p. 115]³¹⁶

Without confidence that our actions make a difference to the world we would not be able to act at all. Therefore, “the openness of

³¹⁶Von Wright argues that because action involves this *counterfactual element*, it is only through action that we can acquire knowledge about causation. For knowing that *C* caused *E*, involves knowing that *E* would not have come about if *C* had not happened. This ‘manipulability theory’, is an important forefather of contemporary interventionist accounts of causation. See, e.g., Pearl [2000] and Woodward [2003].

the future [...] may be said to be baked into our very concept of action" [von Wright 1974, p. 133].

At this point, however, the clever compatibilist may raise an objection. Von Wright's claim is that action is only possible for agents who think that their actions actualise one real possibility from among others. But can agents not *believe* that they are doing so while in fact there is only one future possibility? Hence, this compatibilist will claim that action only requires the future to be *epistemically* open, while it might *ontologically* be determined. In fact, von Wright himself, in the end, adopts this idea that "freedom is relative to human ignorance." So might action then, after all, be possible in a deterministic world?

To me this sort of compatibilism seems very unattractive because it implies that we would lose our ability to act as soon as we learn that the world is deterministic. Action would only be compatible with determinism for as long as we are deceived about the actual deterministic nature of the universe. But let us suppose that we accept this consequence, if only for the sake of argument.³¹⁷ Is this sort of compatibilism untenable? Well, perhaps one might defend it if one adopts an event-causal or non-causal theory of acting.³¹⁸ For on those theories the only thing that makes a happening into an intentional action is that it is causally related to a mental state that represents it, or that one has certain thoughts concurrent to the happening. And the argument in this section might at most establish that one can only think these concurrent thoughts, or be in that mental state, when one believes that the future is open, while it is not required that the

³¹⁷To escape this conclusion a compatibilist might argue that someone who knows that the universe is deterministic still can have the ability to act for as long as he is ignorant about *which* course of action he will take. But I am unconvinced that such an agent can really have the sort of confidence von Wright talks about: the confidence that his actions will make a difference to the world.

³¹⁸Von Wright indeed defends a (sophisticated) form of non-causal action theory. For an excellent discussion of that theory, see Stoutland [1989].

future is actually open for the state to occur, or the thoughts to be thought. However, we have already seen serious reasons to doubt the event-causal and non-causal theories of action in the previous chapters.³¹⁹ In this section we are investigating whether determinism is compatible with the Anscombean understanding of agency. On that understanding, to act is to *know* that you act. And that knowledge is practical, as we saw in the previous chapter, because it makes it the case that what you are doing indeed is an intentional action. This knowledge cannot just be knowledge that our body is moving in a certain way. Knowledge that I am moving my leg when the doctor triggers my knee-reflex, for instance, is no practical knowledge and it does not make my leg movement intentional. Hence in acting we must *know* that we are realising the concept of ‘intentional action’. And since the openness of the future is baked into the very concept of action, practical knowledge is knowledge of that openness.

6.8 Concluding remarks

The aim of this chapter was to investigate how practical knowledge is metaphysically possible. To understand the power of practical knowledge we have looked at what powers in general are. We have seen that both reductive and realist theories of powers have a hard time accounting for cases where a power is triggered but its manifestation is seemingly prevented. I have argued that realists can overcome this problem by understanding manifestations as processes. The identity of processes, I have suggested, is determined by the powers of which they are the manifestations. Once we see that all processes have a specific structure produced by their powers, the idea that practical knowledge determines the form of intentional actions is demystified. Finally we have seen how inde-

³¹⁹See Chapter 3 and section 4.1.

terminism naturally finds a place in a metaphysics of powers and processes. Hence the simple picture of free will is vindicated: by acting intentionally human beings have the power to control which alternative possibilities get actualised.

Thomas Nagel once wrote that many of the alternatives that seem open from an agent's internal perspective would seem closed if we could take up an outer point of view. From the outside it is hard to see how anything can ever be up to the agent, how there is anything left the agent can contribute in the grand scheme of the universe:

Everything I do or that anyone else does is part of a larger course of events that no one "does" but that happens. [Nagel 1986, p. 114]

And indeed if the world just served up a large course of successive events, then there would be nothing for agents to do. But if we conceive of the world not just as one particular thing following another, if we get rid of this final Victorian prejudice, if we start thinking of the world as inhabited by substances that have the power to participate in processes, then, suddenly, there is room for freedom.

Conclusion

When a man cannot choose, he ceases to be a man.

—Anthony Burgess³²⁰

Free will is the capacity human beings have to select and perform one action from a range of alternatives. This thesis started by sketching a very simple picture of this capacity: whenever humans act they actualise one of the real possibilities open to them. Many philosophers feel the need to qualify this picture. They argue, for instance, that only some of our actions are free, or that it is sufficient for free will to believe that the possibilities for action are open to us even if they are not really open. I have attempted to show that the simple picture does not need any qualifications, but can be affirmed as it stands.

In recent years the idea that we have free will has been met with a renewed scepticism based on the discoveries of contemporary brain science. Therefore a first step towards a defence of the simple picture is the rejection of the idea that free will is an illusion exposed by neuroscience. In Chapter 1 we have seen that the results of neuroscientific experiments do not warrant the denial of free will. Free will is typically thought of as the ability to control what one is doing based on reasons. But neuroscientists have

³²⁰ A *Clockwork Orange* Part 2, Chapter 1.

focused on the study of arbitrary selection tasks where no deliberation was allowed. On the basis of the results of these studies, free will cannot be rejected. Furthermore it is impossible, I have argued, to investigate non-arbitrary intentional action by means of the same sorts of experiments neuroscientist use to study random selection. Finally, we have seen that neuro-scientific free will scepticism might be based on conceptual confusion about what the idea that humans have free will entails. Neuro-sceptics think that free action has to start with an almost magical event of conscious origination that is totally uncorrelated to anything prior to that event. But free will requires no such wizardry, it is sufficient for freedom that agents are in *control* of what they do.

The notion of control plays a central role in the contemporary philosophical debate about free will we reviewed in Chapter 2. *Compatibilists* believe that free will is reconcilable with determinism and therefore have to explain how control is possible in a world with a fixed future. Their main adversaries, *libertarians*, argue that we have free will but deny its compatibility with determinism, which makes it their task to explain the possibility of control given indeterminism. In light of this orthogonal opposition, it is striking that most compatibilists and libertarians agree on the nature of the control an agent exhibits in acting intentionally: an agent intentionally controls her action if and only if that action is caused by the right beliefs, desires or intentions. Disagreement between libertarians and compatibilists only arises when they discuss which extra factors are required to turn a merely intentional action into a free action. And hence the free will debate has turned into a game of ‘my account can secure more control than yours’. I have argued that if libertarians take their commitment to indeterminism seriously, they should refrain from playing this game. They should not argue that determinism merely diminishes control, but that control is impossible in a deterministic universe. It

follows, I argued, that the libertarian should reject the distinction between intentional action and free action. In order to uphold the simple picture of free will, the libertarian should undermine the inherently compatibilist idea that control consists in causation by mental states and offer an alternative account of action.

The task of undermining this ‘event-causal theory of action’ was taken up in Chapter 3. According to the causal theory agential control consists in the holding of two different relations between an agent’s mental state and her bodily movement. The state should *cause* the movement and the state should *represent* the movement. The causation explains why the movement comes about, and the representational relation explains why the movement is a rational. I have argued that this theory can never deliver a successful account of agency because the two relations of causation and representation are entirely *independent* of one another, whereas an intentional action happens *because* it is rational.

The 4th chapter began the search for an alternative account of action. Both non-causal and agent-causal theories were discussed, but neither could deliver an account of agency that does not suffer from the separation between causality and rationality. Non-causalists believe that they can fully explain acting for a reason without addressing the question of how an action comes about. But because they do not offer a causal story, they can never explain why an action that an agent has reason to perform actually occurs. Agent-causalists, on the other hand, are able to explain the actual occurrence of an action because they claim that it is caused to exist by an agent. At first this suggestion seems mysterious because causation is typically thought of as a relation that holds between events, but contemporary agent-causalists have done a great deal to argue that objects, or substances, can be causes as well. Hence agent-causation is just a special form of substance-causation, which occurs everywhere in the world. Unfortunately,

agent-causalists have a hard time explaining what is special about agent-causation so that it yields agential control. Some agent-causalists claim that agent causation can yield control because it is the exercise of a rational power, but, I have argued, they cannot satisfactorily explain the rationality of that power—let alone the rationality of the movement it is supposed to produce.

In Chapter 5 I argued that Elizabeth Anscombe's account of action explains how rationality does not stop short of the movement, and can therefore deliver a satisfactory account of agential control. According to Anscombe, the capacity for action is a capacity for a special sort of self knowledge. Whenever an agent is doing something intentionally she has practical knowledge that she is doing that thing. Moreover, practical knowledge is *constitutive* of the rational structure of action. This means that an agent's action is aimed at a certain end, only because this agent knows that she is performing her action in order to attain that end. Hence there is no gap between the agent's knowledge that she is doing something and the occurrence of the act—there is no separation of rationality and causality. This Anscombean view is often rejected out of hand because it suggests that an agent can never be wrong about what she is doing. That is *prima facie* ridiculous because we sometimes think that we are doing something while we in fact are unsuccessful. In such cases it seems that there is a gap between what the agent thinks she is doing and what she is actually doing. I argued that the Anscombean can overcome this problem if she conceives of actions as *ongoing processes*. You can be engaged in a process without ever completing it. Hence, even if you cannot know that your action will be successful—that you will ever get it *done*—you can know that you are *doing* it.

Even if Anscombe is right that the capacity to act is a capacity for practical knowledge, the question remains how we are to understand that capacity, and how it fits into the physical world. These

questions were addressed in the 6th chapter. In order to explain how the capacity for practical knowledge is possible I investigated what the nature of capacities is in general. Our understanding of the metaphysics of capacities, or of powers as they are often called, is hampered by the Humean tendency to think of the world as a mere succession of events. I have argued that we can only arrive at a proper understanding of powers if we understand their manifestations as processes rather than events. The identity of a process is determined by the power to engage in that process. Thus practical knowledge is not unique in that it is constitutive of the structure of action. Finally I have argued that the acceptance of a metaphysics of powers and processes delivers a natural understanding of the open future. Which alternatives are really possible at a certain time and place depends on the powers of the objects located there. The exercise of a power is the actualisation of one of the possibilities.

It turns out that our capacity to act is precisely as the simple picture sketches it. Our action consists in realising one of the really possible alternatives open to us. In this way we help to shape the future. Therefore, we are free.

*Two roads diverged in a yellow wood
And sorry I could not travel both
And be one traveller, long I stood
And looked down one as far as I could
To where it bent in the undergrowth*³²¹

³²¹First verse of Robert Frost's poem 'The road not taken' (1920).

References

Adams, Frederick and Alfred R. Mele

- 1989 “The Role of Intention in Intentional Action”, *Canadian Journal of Philosophy*, 19(4), pp. 511–531. (Cited on pp. 106, 111.)

Anscombe, G.E.M.

- 1957 *Intention*, Harvard University Press, Cambridge, Mass. (Cited on pp. 35, 169, 171–174, 176–180, 186–188, 193, 194, 197, 198, 200, 205–207.)
- 1963 “The Two Kinds of Error in Action”, *The Journal of Philosophy*, 60(14), pp. 393–401. (Cited on p. 200.)
- 1971 “Causality and Determination”, in Anscombe [1981], pp. 133–47. (Cited on p. 223.)
- 1974 “Von Wright on Practical Inference”, in Schilpp *et al.* [1989], pp. 377–404. (Cited on pp. 91, 130, 175.)
- 1975 “The First Person”, in Guttenplan [1975], pp. 45–65. (Cited on p. 198.)
- 1979 “Under a Description”, *Noûs*, 13(2), pp. 219–233. (Cited on p. 172.)
- 1981 *Collected Philosophical Papers, Volume 2: Metaphysics and the Philosophy of Mind*, Wiley-Blackwell, Oxford. (Cited on p. 259.)
- 1982 “Medallist’s Address: Action, Intention and ‘Double Effect’”, *Proceedings of the American Catholic Philosophical Association*, 56, pp. 12–25. (Cited on p. 186.)
- 2005 “Analytical Philosophy and the Spirituality of Man”, in M. Geach *et al.* [2005], pp. 3–16. (Cited on p. 171.)

Antony, Louise

- 1989 “Anomalous Monism and the Problem of Explanatory Force”, *The Philosophical Review*, 98(2), pp. 153–187. (Cited on p. 113.)

Aristotle

- 1996 *Physics*, Translated by Robin Waterfield, Oxford University Press, Oxford. (Cited on p. 177.)
- 1998 *Metaphysics*, Translated by Hugh Lawson-Tancred, Penguin Books, London. (Cited on pp. 3, 177.)

Armstrong, David M.

- 1973 *Belief, Truth and Knowledge*, Cambridge University Press, London. (Cited on p. 239.)
- 1983 *What is a Law of Nature?*, Cambridge University Press, Cambridge. (Cited on p. 149.)

Arpaly, Nomy

- 2006 *Merit, Meaning, and Human Bondage: An Essay on Free Will*, Princeton University Press, Princeton. (Cited on pp. 113, 125.)

Austin, J.L.

- 1952 “How to Talk: Some Simple Ways”, in *Proceedings of the Aristotelian Society*, vol. 53, pp. 227–246. (Cited on p. 193.)

Baer, John, James C. Kaufman and Roy F. Baumeister

- 2008 (eds.), *Are We Free? Psychology and Free Will*, Oxford University Press, New York. (Cited on pp. 260, 286.)

Balaguer, Mark

- 2004 “A Coherent, Naturalistic, and Plausible Formulation of Libertarian Free Will”, *Noûs*, 38(3), pp. 379–406. (Cited on p. 72.)
- 2010 *Free Will as an Open Scientific Problem*, MIT Press, Cambridge, Mass. (Cited on p. 72.)

Banks, William P. and Susan Pockett

- 2007 “Benjamin Libet’s Work on the Neuroscience of Free Will”, in Velmans *et al.* [2007], pp. 657–670. (Cited on p. 27.)

Bargh, John A.

- 2008 “Free Will is Un-natural”, in Baer *et al.* [2008], pp. 128–154. (Cited on p. 55.)

- Bargh, John A., Mark Chen and Lara Burrows
1996 “Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action.” *Journal of personality and social psychology*, 71(2), p. 230. (Cited on p. 55.)
- Bayne, Tim
2011 “Libet and the Case for Free Will Scepticism”, in Swinburne [2011], pp. 25–46. (Cited on pp. 38, 49.)
- Bennett, M.R. and P.M.S. Hacker
2003 *Philosophical Foundations of Neuroscience*, Blackwell Publishing, Oxford. (Cited on p. 23.)
- Bird, Alexander
1998 “Dispositions and Antidotes”, *The Philosophical Quarterly*, 48(191), pp. 227–34. (Cited on p. 218.)
2007 *Nature’s Metaphysics: Laws and Properties*, Oxford University Press, New York. (Cited on p. 219.)
- Bishop, John
1989 *Natural Agency: An Essay on the Causal Theory of Action*, Cambridge University Press, Cambridge. (Cited on pp. 63, 78, 85, 100, 104, 110, 131, 208, 248.)
- Bode, Stefan, Anna Hanxi He, Chun Siong Soon, Robert Trampel, Robert Turner and John-Dylan Haynes
2011 “Tracking the Unconscious Generation of Free Decisions Using Ultra-high Field fMRI”, *PloS one*, 6(6), e21612. (Cited on p. 46.)
- Bonevac, Daniel, Josh Dever and David Sosa
2012 “The Counterexample Fallacy”, *Mind*, 120(480), pp. 1143–1158. (Cited on p. 219.)
- Bourget, David and David J. Chalmers
2014 “What Do Philosophers Believe?”, *Philosophical Studies*, 170(3), pp. 465–500. (Cited on p. 2.)
- Brand, Myles
1989 “Proximate Causation of Action”, *Philosophical Perspectives*, 3, pp. 423–442. (Cited on pp. 104, 108.)
- Bratman, Michael E.
1987 *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, Mass. (Cited on pp. 97, 99, 108.)

Bratman, Michael E.

- 2014 *Shared Agency: A Planning Theory of Acting Together*, Oxford University Press, New York. (Cited on pp. 97, 112.)

Brewer, Bill

- 1995 “Mental Causation: Compulsion by Reason”, *Proceedings of the Aristotelian Society, Supplementary Volume*, 69, pp. 237–253. (Cited on p. 113.)

Broadie, Sarah

- 2013 “Agency and Determinism in A Metaphysics for Freedom”, *Inquiry*, 56(6), pp. 571–582. (Cited on p. 152.)

Buchak, Lara

- 2013 “Free Acts and Chance: Why The Rollback Argument Fails”, *The Philosophical Quarterly*, 63(250), pp. 20–28. (Cited on p. 74.)

Carnap, Rudolf

- 1936 “Testability and Meaning”, *Philosophy of science*, 3(4), pp. 419–471. (Cited on p. 214.)

Cartwright, Nancy and John Pemberton

- 2013 “Aristotelian Powers: Without Them, What Would Modern Science Do?”, in Groff *et al.* [2013], pp. 93–112. (Cited on p. 213.)

Chakravartty, Anjan

- 2007 *A Metaphysics for Scientific Realism: Knowing the Unobservable*, Cambridge University Press, Cambridge. (Cited on p. 220.)

Chisholm, Roderick

- 1966 “Freedom and Action”, in Lehrer [1966], pp. 11–44. (Cited on p. 108.)

Choi, Sungho

- 2005 “Dispositions and Mimickers”, *Philosophical Studies*, 122(2), pp. 183–188. (Cited on p. 217.)
- 2006 “The Simple vs. Reformed Conditional Analysis of Dispositions”, in *Synthese*, 148(2), pp. 369–379. (Cited on p. 217.)
- 2008 “Dispositional Properties and Counterfactual Conditionals”, *Mind*, 117(468), pp. 795–841. (Cited on p. 217.)

Choi, Sungho and Michael Fara

- 2012 “Dispositions”, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Spring 2012. (Cited on p. 232.)

Clancy, Sean

- 2013 “A Strong Compatibilist Account of Settling”, *Inquiry*, 56(6), pp. 653–665. (Cited on p. 152.)

Clarke, Randolph

- 1993 “Toward a Credible Agent-causal Account of Free Will”, *Noûs*, 27(2), pp. 191–203. (Cited on p. 149.)
- 2003 *Libertarian Accounts of Free Will*, Oxford University Press, New York. (Cited on pp. 12, 72, 80, 145, 148, 149, 156, 157.)

Clarke, Randolph and Justin Capes

- 2014 “Incompatibilist (Nondeterministic) Theories of Free Will”, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Spring 2014. (Cited on p. 145.)

Cohen, S. Marc

- 2014 “Aristotle’s Metaphysics”, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Spring 2014. (Cited on p. 3.)

Cross, Troy

- 2012 “Recent Work on Dispositions”, *Analysis*, 72(1), pp. 115–124. (Cited on p. 220.)

Davidson, Donald

- 1963 “Actions, Reasons, and Causes”, *The Journal of Philosophy*, 60(23), pp. 685–700. (Cited on pp. 15, 60, 70, 94, 139, 170, 176.)
- 1967 “Truth and Meaning”, *Synthese*, 17(1), pp. 304–323. (Cited on p. 127.)
- 1970 “Mental Events”, in Davidson [2001], ch. 11, pp. 207–227. (Cited on pp. 96, 130, 142.)
- 1973 “Freedom to Act”, in Davidson [2001], ch. 4, pp. 63–82. (Cited on pp. 88, 102, 109.)
- 1978 “Intending”, in Davidson [2001], ch. 6, pp. 83–102. (Cited on pp. 97, 103, 127, 201.)
- 1982 “Paradoxes of Irrationality”, in Davidson [2004], ch. 11, pp. 169–188. (Cited on pp. 96, 128.)

Davidson, Donald

- 1990 "Representation and Interpretation", in Davidson [2004], pp. 87–99. (Cited on p. 119.)
- 2001 *Essays on Actions and Events (2nd edition)*, Clarendon Press, Oxford. (Cited on pp. 94, 263.)
- 2004 *Problems of Rationality*, Oxford University Press, Oxford. (Cited on pp. 123, 263, 264.)

De Caro, Mario and David Macarthur

- 2004 *Naturalism in Question*, Harvard University Press, Cambridge, Mass. (Cited on p. 269.)

Dennett, Daniel C.

- 1981 "True Believers: The Intentional Strategy and Why it Works", in Heath [1981], pp. 53–75. (Cited on p. 48.)
- 2004 *Freedom Evolves*, Penguin Books, London. (Cited on pp. 36, 59, 61.)

Dennett, Daniel C. and Marcel Kinsbourne

- 1992 "Time and the Observer: The Where and When of Consciousness in the Brain", *Behavioral and Brain Sciences*, 15(2), pp. 183–201. (Cited on p. 25.)

Diamond, Cora and Jenny Teichman

- 1979 (eds.), *Intention and Intentionality: Essays in Honour of G.E.M. Anscombe*, Cornell University Press, Ithaca, New York. (Cited on p. 277.)

Dowty, David R.

- 1977 "Toward a Semantic Analysis of Verb Aspect and the English 'Imperfective' Progressive", *Linguistics and Philosophy*, 1(1), pp. 45–77. (Cited on p. 200.)

Dretske, Fred

- 1989 "Reasons and Causes", *Philosophical Perspectives*, 3, pp. 1–15. (Cited on p. 97.)

Eccles, John C.

- 1985 "Mental Summation: The Timing of Voluntary Intentions by Cortical Activity", *Behavioral and Brain Sciences*, 8(4), pp. 542–543. (Cited on p. 52.)

Edwards, Jonathan

- 1754/2009 *Freedom of the Will, The Works of Jonathan Edward, Vol. I*, ed. by Paul Ramsey, Yale University Press, New Haven Conn. (Cited on pp. 242, 243.)

Ekstrom, Laura Waddell

- 2000 *Free Will: A Philosophical Study*, Westview Press, Boulder, Col. (Cited on p. 12.)

Ellis, Brian

- 2001 *Scientific Essentialism*, Cambridge University Press, Cambridge. (Cited on pp. 220, 222, 229.)

Enç, Berent

- 2003 *How We Act: Causes, Reasons, and Intentions*, Oxford University Press, New York. (Cited on p. 122.)
- 2004 “Causal Theories of Intentional Behavior and Wayward Causal Chains”, *Behavior and Philosophy*, 32(1), pp. 149–166. (Cited on pp. 116, 117.)

Falvey, Kevin

- 2000 “Knowledge in Intention”, *Philosophical Studies*, 99(1), pp. 21–44. (Cited on pp. 170, 173, 187.)

Fara, Michael

- 2005 “Dispositions and Habituals”, *Noûs*, 39(1), pp. 43–82. (Cited on pp. 217, 219.)

Feigl, Herbert, Michael Scriven and Grover Maxwell

- 1958 (eds.), *Concepts, Theories, and the Mind-Body Problem*, University of Minnesota Press, Minneapolis. (Cited on p. 282.)

Filevich, Elisa, Simone Kühn and Patrick Haggard

- 2013 “There is no Free Won’t: Antecedent Brain Activity Predicts Decisions to Inhibit”, *PLoS one*, 8(2), e53053. (Cited on p. 28.)

Fischer, John Martin, Robert Kane, Derk Pereboom and Manuel Vargas

- 2007 *Four Views on Free Will*, Blackwell Publishing, Oxford. (Cited on pp. 89, 279.)

Fischer, John Martin and Mark Ravizza

- 2000 *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, New York. (Cited on pp. 14, 69, 84.)

Flanagan, Owen J.

- 1992 *Consciousness Reconsidered*, MIT press, Cambridge, Mass. (Cited on p. 33.)

Fodor, Jerry A.

- 1992 “Making Mind Matter More”, *Philosophical Topics*, 17(1), pp. 59–79. (Cited on p. 97.)

Ford, Anton, Jennifer Hornsby and Frederick Stoutland

- 2011 (eds.), *Essays on Anscombe’s Intention*, Harvard University Press, Cambridge, Mass. (Cited on pp. 170, 274, 277, 280, 282, 285.)

Frankfurt, Harry G.

- 1969 “Alternate Possibilities and Moral Responsibility”, *Journal of Philosophy*, 66(3), pp. 829–39. (Cited on pp. 9, 84.)
1978 “The Problem of Action”, *American Philosophical Quarterly*, 15(2), pp. 157–162. (Cited on p. 105.)

Franklin, Christopher Evan

- 2011a “Farewell to the Luck (and Mind) Argument”, *Philosophical Studies*, 156(2), pp. 199–230. (Cited on pp. 71, 72, 74, 75, 80, 82.)
2011b “The Problem of Enhanced Control”, *Australasian Journal of Philosophy*, 89(4), pp. 687–706. (Cited on pp. 79, 81, 82.)
2014 “Event-Causal Libertarianism, Functional Reduction, and the Disappearing Agent Argument”, *Philosophical Studies*, 170(3), pp. 413–432. (Cited on pp. 72, 99.)

Fried, Itzhak, Amiram Katz, Gregory McCarthy, Kimberlee J. Sass, Peter Williamson, Susan S. Spencer and Dennis D. Spencer

- 1991 “Functional Organization of Human Supplementary Motor Cortex Studied by Electrical Stimulation”, *The Journal of Neuroscience*, 11(11), pp. 3656–3666. (Cited on p. 38.)

Gazzaniga, Michael

- 2012 *Who’s in Charge?: Free Will and the Science of the Brain*, Constable & Robinson Ltd, London. (Cited on p. 20.)

Geach, Mary and Luke Gormally

- 2005 (eds.), *Human Life, Action and Ethics: Essays by G.E.M. Anscombe*, Imprint Academic, Exeter. (Cited on pp. 3, 170, 259.)

Geach, Peter T.

- 1972 *Logic Matters*, University of California Press, Berkeley. (Cited on p. 228.)

Gibb, Sophie, E.J. Lowe and R.D. Ingthorsson

- 2013 (eds.), *Mental Causation and Ontology*, Oxford University Press, Oxford. (Cited on p. 270.)

Ginet, Carl

- 1990 *On Action*, Cambridge University Press, New York. (Cited on p. 145.)
- 1997 “Freedom, Responsibility, and Agency”, *Journal of Ethics*, 1(1), pp. 85–98. (Cited on p. 145.)
- 2000 “Book Review. The Works of Agency: On Human Action, Will, and Freedom. Hugh McCann.” *Philosophical Review*, 109(4), pp. 632–635. (Cited on p. 145.)
- 2008 “In Defense of a Non-causal Account of Reasons Explanations”, *The Journal of Ethics*, 12(3), pp. 229–237. (Cited on pp. 145, 146.)

Glock, Hans-Johann

- 2014 “Reasons for Action: Wittgensteinian and Davidsonian Perspectives in Historical, Meta-philosophical and Philosophical Context”, *Nordic Wittgenstein Review*, 3(1), pp. 7–46. (Cited on p. 98.)

Goetz, Stewart C.

- 1988 “A Noncausal Theory of Agency”, *Philosophy and Phenomenological Research*, 49(2), pages. (Cited on p. 144.)

Goldman, Alvin I.

- 1970 *A Theory of Human Action*, Princeton University Press, Princeton. (Cited on p. 124.)
- 1979 “Action, Causation, and Unity”, *Noûs*, 13(2), pp. 261–270. (Cited on p. 97.)

Goodman, Nelson

- 1955 *Fact, Fiction and Forecast*, Harvard University Press, Cambridge, Mass. (Cited on pp. 215, 216.)

Griffith, Meghan

- 2010 “Why Agent-caused Actions are Not Lucky”, *American Philosophical Quarterly*, 47(1), pp. 43–56. (Cited on p. 101.)

Groff, Ruth

- n.d. "Sublating the Free Will Problematic: Powers, Agency and Causal Determination", Unpublished manuscript. (Cited on pp. 150, 151, 246, 247.)
- 2013 *Ontology Revisited: Metaphysics in Social and Political Philosophy*, Routledge, New York. (Cited on p. 220.)

Groff, Ruth and John Greco

- 2013 (eds.), *Powers and Capacities in Philosophy: The New Aristotelianism*, Routledge, New York. (Cited on pp. 262, 273.)

Grünbaum, Thor

- 2009 "Anscombe and Practical Knowledge of What is Happening", *Grazer Philosophische Studien*, 78, pp. 41–67. (Cited on pp. 187, 189, 195.)
- 2012 "Seeing what I am doing", *Philosophy and Phenomenological Research*, 86(2). (Cited on p. 207.)

Gundersen, Lars

- 2002 "In Defence of the Conditional Account of Dispositions", *Synthese*, 130(3), pp. 389–411. (Cited on p. 217.)

Guttenplan, Samuel

- 1975 (ed.), *Mind and Language: Wolfson College Lectures 1974*, Clarendon Press, Oxford. (Cited on p. 259.)

Haddock, Adrian

- 2010 "Knowledge and Action", in Pritchard *et al.* [2010], pp. 244–328. (Cited on pp. 193, 195.)

Haggard, Patrick and Martin Eimer

- 1999 "On the Relation Between Brain Potentials and the Awareness of Voluntary Movements", *Experimental Brain Research*, 126(1), pp. 128–133. (Cited on pp. 27, 32, 38, 40.)

Haji, Ishtiyaque

- 1999 "Indeterminism and Frankfurt-type Examples", *Philosophical Explorations*, 2(1), pp. 42–58. (Cited on p. 75.)
- 2000 "Indeterminism, Explanation, and Luck", *Journal of Ethics*, 4(3), pp. 211–235. (Cited on p. 75.)

Handfield, Toby

- 2009 *Dispositions and Causes*, Oxford University Press, Oxford. (Cited on p. 278.)

Harman, Gilbert

- 1976 “Practical Reasoning”, *The Review of Metaphysics*, 29(3), pp. 431–463. (Cited on p. 108.)

Harré, R. and E.H. Madden

- 1975 *Causal Powers: A Theory of Natural Necessity*, Basil Blackwell, Oxford. (Cited on p. 220.)

Harris, Sam

- 2012 *Free Will*, Simon and Schuster, New York. (Cited on p. 20.)

Haynes, John-Dylan

- 2011 “Decoding and Predicting Intentions”, *Annals of the New York Academy of Sciences*, 1224(1), pp. 9–21. (Cited on pp. 40, 43, 44, 46, 48, 49, 51, 66.)

Heath, A. F.

- 1981 (ed.), *Scientific Explanation: Papers Based on Herbert Spencer Lectures Given in the University of Oxford*, Clarendon Press, Oxford. (Cited on p. 264.)

Heil, John

- 2004 “Review of Powers: A Study in Metaphysics by George Molnar”, *Journal of Philosophy*, 101(8), pp. 438–443. (Cited on p. 235.)
- 2005 “Dispositions”, *Synthese*, 144(3), pp. 343–56. (Cited on p. 220.)
- 2012 *The Universe As We Find It*, Oxford University Press, Oxford. (Cited on pp. 11, 224.)

Hobbes, Thomas

- 1651/2012 *Leviathan*, Dover Publications, Mineola, NY. (Cited on p. 87.)

Hornsby, Jennifer

- 2004a “Agency and Actions”, *Royal Institute of Philosophy Supplement*, 55, pp. 1–23. (Cited on p. 99.)
- 2004b “Agency and Alienation”, in de Caro *et al.* [2004], pp. 173–187. (Cited on p. 99.)
- 2012 “Actions and Activity”, *Philosophical Issues*, 22(1), pp. 233–245. (Cited on pp. 201, 226.)

Horst, David

- 2012 *Absichtliches Handeln*, mentis Verlag, Paderborn. (Cited on pp. 128, 170.)

Hume, David

- 1748/2011 *An Enquiry Concerning Human Understanding*, ed. by Lorne Falkenstein, Broadview Press, Claremont. (Cited on p. 214.)

Hutchison, Keith

- 1991 “Dormitive Virtues, Scholastic Qualities, and the New Philosophies”, *History of Science*, 29, pp. 245–278. (Cited on p. 240.)

Hüttemann, Andreas

- 2013 “A Disposition-Based Process Theory of Causation”, in Mumford *et al.* [2013], ch. 5, pp. 101–22. (Cited on p. 220.)

Jacobs, Jonathan D. and Timothy O’Connor

- 2013 “Agent Causation in a Neo-Aristotelian Metaphysics”, in Gibb *et al.* [2013], pp. 173–192. (Cited on pp. 83, 157, 161.)

Johnston, Mark

- 1992 “How to Speak of the Colors”, *Philosophical Studies*, 68(3), pp. 221–263. (Cited on p. 218.)

Kalis, Annemarie, Stefan Kaiser and Andreas Mojzisch

- 2013 “Why We Should Talk About Option Generation In Decision-making Research”, *Frontiers in psychology*, 4(555). (Cited on p. 46.)

Kane, Robert

- 1985 *Free Will and Values*, State University of New York Press, New York. (Cited on pp. 10, 72.)
- 1996 *The Significance of Free Will*, Oxford University Press, New York. (Cited on pp. 10, 72.)
- 1999a “On Free Will, Responsibility and Indeterminism: Responses to Clarke, Haji, and Mele”, *Philosophical Explorations*, 2, pp. 105–121. (Cited on pp. 30, 72.)
- 1999b “Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism”, *Journal of Philosophy*, 96(5), pp. 217–240. (Cited on pp. 10, 14, 72.)
- 2002 (ed.), *The Oxford Handbook of Free will*, Oxford University Press, Oxford. (Cited on pp. 278, 279.)
- 2011 *The Oxford Handbook of Free Will: Second Edition*, Oxford University Press, New York. (Cited on p. 278.)

Keil, Geert

- 2007 “What Do Deviant Causal Chains Deviate From?”, in Lumer *et al.* [2012], pp. 69–90. (Cited on pp. 109, 124.)

Keller, Ivonne and Heinz Heckhausen

- 1990 “Readiness Potentials Preceding Spontaneous Motor Acts: Voluntary vs. Involuntary Control”, *Electroencephalography and clinical Neurophysiology*, 76(4), pp. 351–361. (Cited on pp. 27, 38.)

Kenny, Anthony

- 1963 *Action, Emotion and Will*, Routledge & Kegan Paul, London. (Cited on p. 94.)

Kim, Jaegwon

- 2007 “Causation and Mental Causation”, in McLaughlin *et al.* [2007], pp. 227–242. (Cited on p. 97.)

Kim, Seong-Gi, Wolfgang Richter and Kāmil Uğurbil

- 1997 “Limitations of Temporal Resolution in Functional MRI”, *Magnetic Resonance in Medicine*, 37(4), pp. 631–636. (Cited on p. 42.)

Kistler, Max and Bruno Gnessounou

- 2012 *Dispositions and Causal Powers*, Ashgate Publishing Ltd., Farnham. (Cited on p. 276.)

Klein, Stanley A.

- 2002a “Libet’s Research on the Timing of Conscious Intention to Act: A Commentary”, *Consciousness and Cognition*, 11(2), pp. 273–279. (Cited on p. 27.)
- 2002b “Libet’s Temporal Anomalies: A Reassessment of the Data”, *Consciousness and Cognition*, 11(2), pp. 198–214. (Cited on p. 27.)
- 2002c “Libet’s Timing of Mental Events: Commentary on the Commentaries”, *Consciousness and Cognition*, 11(2), pp. 326–333. (Cited on p. 27.)

Klemm, W.R.

- 2010 “Free Will Debates: Simple Experiments are Not so Simple”, *Advances in Cognitive Psychology*, 6, p. 47. (Cited on p. 20.)

Kornhuber, Hans H. and Lüder Deecke

- 1965 “Hirnpotentialänderungen bei Willkürbewegungen und Passiven Bewegungen des Menschen: Bereitschaftspotential und Reafferente Potentiale”, *Pflüger’s Archiv für die Gesamte Physiologie des Menschen und der Tiere*, 284(1), pp. 1–17. (Cited on p. 21.)

Kripke, Saul A.

- 1972 *Naming and Necessity*, Harvard University Press, Cambridge, Mass. (Cited on p. 221.)

Lamme, Victor

- 2011a *De Vrije Wil Bestaat Niet*, Prometheus, Amsterdam. (Cited on pp. 20, 56.)
2011b “Vrij Amateurstisch”, *NRC Next*, (20-09-2011). (Cited on p. 39.)

Laplace, Pierre Simon Marquis De

- 1814/2007 *A Philosophical Essay on Probabilities*, Translated by F.W. Truscott and F.L. Emory, Cosimo Inc., New York. (Cited on p. 66.)

Lau, Hakwan C., Robert D. Rogers, Patrick Haggard and Richard E. Passingham

- 2004 “Attention to Intention”, *Science*, 303(5661), pp. 1208–1210. (Cited on p. 25.)

Lau, Hakwan C., Robert D. Rogers and Richard E. Passingham

- 2006 “On Measuring the Perceived Onsets of Spontaneous Actions”, *The Journal of Neuroscience*, 26(27), pp. 7265–7271. (Cited on p. 26.)

Lavin, Douglas

- 2013 “Must There Be Basic Action?”, *Noûs*, 47(2), pp. 273–301. (Cited on pp. 97, 170, 192.)

Lehrer, Keith

- 1966 (ed.), *Freedom and Determinism*, Random House, New York. (Cited on p. 262.)

LePore, Ernest and Brian P. McLaughlin

- 1985 (eds.), *Actions and Events; Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, New York. (Cited on pp. 228, 287.)

Lewis, David K.

- 1973 *Counterfactuals*, Blackwell Publishing, Oxford. (Cited on p. 217.)
- 1979 “Counterfactual Dependence and Time’s Arrow”, *Noûs*, 13(4), pp. 455–76. (Cited on p. 244.)
- 1980 “Veridical Hallucination and Prosthetic Vision”, *Australasian Journal of Philosophy*, 58(3), pp. 239–249. (Cited on p. 124.)
- 1986a *On the Plurality of Worlds*, Blackwell Publishing, Oxford. (Cited on p. 215.)
- 1986b *Philosophical Papers: Volume II*, Oxford University Press, New York. (Cited on pp. 215, 224, 226.)
- 1997 “Finkish Dispositions”, *The Philosophical Quarterly*, 47(187), pp. 143–58. (Cited on pp. 217, 218, 232.)

Libet, Benjamin

- 1985 “Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action”, *Behavioral and Brain Sciences*, 8(4), pp. 529–539. (Cited on pp. 14, 19, 22, 23.)

Libet, Benjamin, Curtis A. Gleason, Elwood W. Wright and Dennis K. Pearl

- 1983 “Time of Conscious Intention to Act in Relation to Onset of Cerebral Activity (Readiness-Potential) The Unconscious Initiation of a Freely Voluntary Act”, *Brain*, 106(3), pp. 623–642. (Cited on pp. 14, 19, 22, 26–28, 32, 34.)

Lowe, E.J.

- 2008 *Personal Agency: The Metaphysics of Mind and Action*, Oxford University Press, Oxford. (Cited on pp. 149, 163, 164.)
- 2013 “The Will as a Rational Free Power”, in Groff *et al.* [2013], pp. 172–185. (Cited on p. 163.)

Lumer, Christoph and Sandro Nannini

- 2012 *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*, Ashgate Publishing Ltd., Farnham. (Cited on pp. 271, 275.)

Mackie, J. L.

- 1977 “Dispositions, Grounds, and Causes”, *Synthese*, 34(4), pp. 361–369. (Cited on p. 239.)

Macrae, C. Neil and Lucy Johnston

- 1998 “Help, I Need Somebody: Automatic Action and Inaction”, *Social Cognition*, 16(4), pp. 400–417. (Cited on p. 56.)

Malzkorn, Wolfgang

- 2000 “Realism, Functionalism and the Conditional Analysis of Dispositions”, *Philosophical Quarterly*, 50(201), pp. 452–469. (Cited on p. 219.)

Manley, David and Ryan Wasserman

- 2008 “On Linking Dispositions and Conditionals”, *Mind*, 117(465), pp. 59–84. (Cited on pp. 217, 219.)

Markosian, Ned

- 1999 “A Compatibilist Version of the Theory of Agent Causation”, *Pacific Philosophical Quarterly*, 80(3), pp. 257–277. (Cited on p. 148.)
- 2012 “Agent Causation as the Solution to all the Compatibilist’s Problems”, *Philosophical Studies*, 157(3), pp. 383–398. (Cited on p. 148.)

Marmodoro, Anna

- 2009 “Do Powers Need Powers to Make them Powerful?: From Pandispositionalism to Aristotle”, *History of Philosophy Quarterly*, 26(4). (Cited on p. 220.)
- 2010 (ed.), *The Metaphysics of Powers: Their Grounding and their Manifestations*, Routledge, New York. (Cited on p. 220.)

Martin, C.B.

- 1994 “Dispositions and Conditionals”, *The Philosophical Quarterly*, 44(174), pp. 1–8. (Cited on pp. 217, 231.)
- 2007 *The Mind in Nature*, Oxford University Press, New York. (Cited on pp. 220, 224.)

Mayr, Erasmus

- 2011 *Understanding Human Agency*, Oxford University Press, New York. (Cited on pp. 107, 111, 117, 122, 161.)

McCann, Hugh

- 1998 *The Works of Agency: On Human Action, Will, and Freedom*, Cornell University Press, New York. (Cited on pp. 144, 145.)

McDowell, John

- 2011 “Anscombe on Bodily Self-Knowledge”, in Ford *et al.* [2011], pp. 128–146. (Cited on p. 174.)

McGinn, Colin

- 1989 “Can We Solve the Mind-Body Problem?”, *Mind*, 98(391), pp. 349–366. (Cited on p. 69.)

McGuire, John Michael

- 2007 “Actions, Reasons, and Intentions: Overcoming Davidson’s Ontological Prejudice”, *Dialogue*, 46(3), pp. 459–479. (Cited on p. 92.)

McKenna, Michael

- 2009 “Compatibilism: State of the Art”, in *The Stanford Encyclopedia of Philosophy*, ed. by Edward N. Zalta, Winter 2009. (Cited on p. 69.)

McLaughlin, Brian P. and Jonathan Cohen

- 2007 (eds.), *Contemporary Debates in Philosophy of Mind*, Blackwell Publishing, Oxford. (Cited on p. 271.)

McTaggart, John M. E.

- 1927 *The Nature of Existence – Volume II*, Cambridge University Press, Cambridge. (Cited on p. 228.)

Melden, A.I.

- 1967 *Free Action*, Routledge & Kegan Paul, London. (Cited on pp. 94, 135.)

Mele, Alfred R.

- 1992 *Springs of Action: Understanding Intentional Behavior*, Oxford University Press, New York. (Cited on pp. 104–106.)
- 1995 *Autonomous Agents: From Self-control to Autonomy*, Oxford University Press, Oxford. (Cited on p. 12.)
- 2006 *Free Will and Luck*, Oxford University Press, New York. (Cited on pp. 68, 80, 81.)
- 2007 “Free Will: Action Theory Meets Neuroscience”, in Lumer *et al.* [2012]. (Cited on p. 37.)
- 2009 *Effective Intentions: The Power of Conscious Will*, Oxford University Press, New York. (Cited on p. 37.)
- 2014 “Free Will and Substance Dualism: The Real Scientific Threat to Free Will?”, in Sinnott-Armstrong [2014], pp. 195–208. (Cited on pp. 39, 45, 59.)

Mellor, D.H.

- 1974 “In Defense of Dispositions”, *Philosophical Review*, 83(2), pp. 157–181. (Cited on p. 211.)

Mellor, D.H.

- 1981 *Real Time*, Cambridge University Press, Cambridge. (Cited on p. 228.)
- 2000 “The Semantics and Ontology of Dispositions”, *Mind*, 109(436), pp. 757–780. (Cited on p. 219.)

Michon, Cyrille

- 2007 “Opium’s Virtus Dormitiva”, in Kistler *et al.* [2012], pp. 133–150. (Cited on p. 239.)

Millgram, Elijah

- 2001 (ed.), MIT Press. (Cited on p. 286.)

Millican, Peter

- 2011 “Hume’s Determinism”, *Canadian Journal of Philosophy*, 40(4), pp. 611–642. (Cited on p. 244.)

Millikan, Ruth Garrett

- 1984 *Language, Thought, and Other Biological Categories*, The MIT Press, Cambridge, Mass. (Cited on p. 116.)
- 1989 “In Defense of Proper Functions”, *Philosophy of Science*, 56(2), pp. 288–302. (Cited on p. 116.)
- 1993 *White Queen Psychology and Other Essays for Alice*, The MIT Press, Cambridge, Mass. (Cited on p. 92.)
- 1995 “Pushmi-pullyu Representations”, *Philosophical perspectives*, 9, pp. 185–200. (Cited on p. 198.)

Molnar, George

- 2003 *Powers: A Study in Metaphysics*, Oxford University Press, New York. (Cited on pp. 220, 235.)

Montague, P. Read

- 2008 “Free Will”, *Current biology*, 18(14), pp. 584–585. (Cited on p. 59.)

Moore, James W., Daniel M. Wegner and Patrick Haggard

- 2009 “Modulating the Sense of Agency with External Cues”, *Consciousness and Cognition*, 18(4), pp. 1056–1064. (Cited on p. 26.)

Moran, Richard

- 2004 “Anscombe on ‘Practical Knowledge’”, *Royal Institute of Philosophy Supplement*, 55, pp. 43–68. (Cited on pp. 170, 177, 187, 189, 190, 193, 194.)

Moran, Richard and Martin J. Stone

- 2011 “Anscombe on Expression of Intention”, in Ford *et al.* [2011].
(Cited on pp. 171, 192, 196.)

Mulder, Jesse M.

- 2014 *Conceptual Realism; The Structure of Metaphysical Thought*,
Quaestiones Infnitae, Volume 81, Utrecht. (Cited on pp. 6,
237.)

Müller, Anselm Winfried

- 1979 “How Theoretical is Practical Reason?”, in Diamond *et al.*
[1979], pp. 91–108. (Cited on p. 175.)
1999 “Was Heisst ”Praktisches Wissen“?”, *Zeitschrift für Philo-*
sophische Forschung, 45, pp. 545–557. (Cited on p. 170.)

Mumford, Stephen

- 1998 *Dispositions*, Oxford University Press, New York. (Cited on
pp. 220, 221, 229, 239.)
2001 “Realism and the Conditional Analysis of Dispositions: Reply
to Malzkorn”, *Philosophical Quarterly*, 51(204), pp. 375–378.
(Cited on p. 220.)
2004 *Laws in Nature*, Routledge, New York. (Cited on p. 220.)

Mumford, Stephen and Rani Lill Anjum

- 2011 *Getting Causes from Powers*, Oxford University Press, New
York. (Cited on pp. 223, 224.)

Mumford, Stephen and Matthew Tugby

- 2013 (eds.), *Metaphysics and Science*, Oxford University Press,
New York. (Cited on p. 270.)

Murphy, Nancey, George F.R. Ellis and Timothy O’Connor

- 2009 (eds.), *Downward Causation and the Neurobiology of Free
Will*, Springer, Berlin. (Cited on p. 278.)

Nagel, Thomas

- 1986 *The View From Nowhere*, Oxford University Press, New York.
(Cited on pp. 148, 252.)

Nahmias, Eddy A.

- 2002 “When Consciousness Matters: A Critical Review of Daniel
Wegner’s *The Illusion of Conscious Will*”, *Philosophical Psy-*
chology, 15(4), pp. 527–541. (Cited on p. 57.)

Nelkin, Dana Kay

- 2011 *Making Sense of Freedom and Responsibility*, Oxford University Press, Oxford. (Cited on p. 148.)

Nichols, Shaun and Joshua Knobe

- 2007 “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions”, *Noûs*, 41(4), pp. 663–685. (Cited on p. 2.)

Nisbett, Richard E. and Timothy D. Wilson

- 1977 “Telling More Than We Can Know: Verbal Reports on Mental Processes”, *Psychological Review*, 84(3), p. 231. (Cited on p. 56.)

Nozick, Robert

- 1981 *Philosophical Explanations*, Harvard University Press, Cambridge, Mass. (Cited on p. 11.)

Nucci, Ezio Di

- 2012 “Priming Effects and Free Will”, *International Journal of Philosophical Studies*, 20(5), pp. 725–734. (Cited on p. 55.)

O’Connor, Timothy

- 1995a “Agent Causation”, in O’Connor [1995b], pp. 61–79. (Cited on pp. 149, 159.)
- 1995b (ed.), *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, Oxford University Press, New York. (Cited on p. 278.)
- 2000 *Persons and Causes: The Metaphysics of Free Will*, Oxford University Press, New York. (Cited on pp. 67, 149, 158, 160.)
- 2002 “Libertarian Views: Dualist and Agent-causal Theories”, in Kane [2002], ch. 15, pp. 337–356. (Cited on p. 158.)
- 2009a “Agent-causal Power”, in Handfield [2009]. (Cited on pp. 101, 149, 160.)
- 2009b “Conscious Willing and the Emerging Sciences of Brain and Behavior”, in Murphy *et al.* [2009], pp. 173–186. (Cited on p. 30.)
- 2011 “Agent-causal Theories of Freedom”, in Kane [2011], pp. 309–328. (Cited on pp. 151, 160.)

O’Connor, Timothy and Constantine Sandis

- 2010 (eds.), *A Companion to the Philosophy of Action*, John Wiley & Sons, New York. (Cited on p. 284.)

Paul, Sarah K.

- 2009 “How We Know What We’re Doing”, *Philosophers’ Imprint*, 9(11). (Cited on p. 173.)
- 2011 “Deviant Formal Causation”, *Journal of Ethics and Social Philosophy*, 5(3). (Cited on pp. 181–183, 185.)

Peacocke, Christopher

- 1979 “Deviant Causal Chains”, *Midwest Studies in Philosophy*, 4(1), pp. 123–155. (Cited on pp. 110, 111.)

Pearl, Judea

- 2000 *Causality: Models, Reasoning, and Inference*, New York, Cambridge University Press. (Cited on p. 249.)

Pedersen, Hans and Megan Altman

- 2015 (eds.), *Horizons of Authenticity in Phenomenology, Existentialism, and Moral Psychology: Essays in Honor of Charles Guignon*, Springer, Dordrecht. (Cited on p. 279.)

Pereboom, Derk

- 2002 “Living Without Free Will: the Case for Hard Incompatibilism”, in Kane [2002], pp. 477–88. (Cited on p. 69.)
- 2004 “Is Our Conception of Agent-causation Coherent?”, *Philosophical Topics*, 32(1), pp. 275–286. (Cited on p. 101.)
- 2007 “Hard Incompatibilism”, in Fischer *et al.* [2007], ch. 3, pp. 85–125. (Cited on pp. 69, 101.)
- 2014 *Free Will, Agency, and Meaning in Life*, Oxford University Press, Oxford. (Cited on p. 160.)
- 2015 “The Phenomenology of Agency and Deterministic Agent Causation”, in Pedersen *et al.* [2015], pp. 277–294. (Cited on pp. 101, 152.)

Pickard, Hanna

- 2004 “Knowledge of Action Without Observation”, *Proceedings of the Aristotelian Society*, 104, pp. 205–230. (Cited on p. 173.)

Prior, Arthur N.

- 1962 “Limited Indeterminism”, *The Review of Metaphysics*, 16(1), pp. 55–61. (Cited on pp. 242, 245.)

Pritchard, Duncan, Adrian Haddock and Alan Millar

- 2010 (eds.), *The Nature and Value of Knowledge: Three Investigations*, Oxford University Press, New York. (Cited on p. 268.)

Psillos, Stathis

- 2002 *Causation and Explanation*, Acumen, Chesham. (Cited on p. 221.)

Quine, Willard Van Orman

- 1960 *Word and Object*, MIT Press, Cambridge, Mass. (Cited on p. 216.)

Raz, Joseph

- 2009 “Reasons: Explanatory and Normative”, in Sandis [2009]. (Cited on p. 106.)

Reid, Thomas

- 1895/1983 *The Works of Thomas Reid*, ed. by Sir William Hamilton, Georg Olms Verlag, Hildesheim. (Cited on p. 148.)

Rödl, Sebastian

- 2007 *Self-Consciousness*, Harvard University Press, Cambridge, Mass. (Cited on pp. 16, 167, 170, 192, 199.)
- 2011 “Two Forms of Practical Knowledge and Their Unity”, in Ford *et al.* [2011], pp. 211–241. (Cited on pp. 175, 177.)
- 2012/2005 *Categories of the Temporal: an Inquiry Into the Forms of the Finite Intellect*, Translated from the German by Sibylle Salewski, Harvard University Press, Cambridge, Mass. (Cited on p. 234.)
- 2014 “Intentional Transaction”, *Philosophical Explorations*, 17(3), pp. 304–316. (Cited on p. 199.)

Roskies, Adina L.

- 2010 “How Does Neuroscience Affect our Conception of Volition?”, *Annual Review of Neuroscience*, 33, pp. 109–130. (Cited on p. 47.)
- 2011 “Why Libet’s Studies Don’t Pose a Threat to Free Will”, in Sinnott-Armstrong *et al.* [2010], pp. 11–22. (Cited on pp. 34, 51, 52.)

Roth, Gerhard

- 2003 *Fühlen, Denken, Handeln: Wie das Gehirn unser Verhalten Steuert*, Suhrkamp, Frankfurt am Main. (Cited on p. 20.)

Rumberg, Antje

- n.d. *Transitions Towards a Semantics for Real Possibility*, Manuscript, PhD thesis. (Cited on p. 246.)

Ryle, Gilbert

1949 *The Concept of Mind*, Hutchinson, London. (Cited on p. 216.)

Sandis, Constantine

2009 (ed.), *New Essays on the Explanation of Action*, Palgrave Macmillan, London. (Cited on pp. 280, 284.)

Sarrazin, Jean-Christophe, Axel Cleeremans and Patrick Haggard

2008 “How Do We Know What we are Doing? Time, Intention and Awareness of Action”, *Consciousness and Cognition*, 17(3), pp. 602–615. (Cited on p. 26.)

Schilpp, Paul Arthur and Lewis Edwin Hahn

1989 (eds.), *The Philosophy of G. H. von Wright*, Open Court, La Salle, Ill. (Cited on pp. 259, 284.)

Schlosser, Markus E.

2007 “Basic Deviance Reconsidered”, *Analysis*, 67(295), pp. 186–194. (Cited on p. 113.)

2012 “Free Will and the Unconscious Precursors of Choice”, *Philosophical Psychology*, 25(3), pp. 365–384. (Cited on pp. 24, 44.)

2014 “The Luck Argument Against Event-causal Libertarianism: It is Here to Stay”, *Philosophical Studies*, 167(2), pp. 375–385. (Cited on pp. 77, 101.)

Schrenk, Markus

2010 “The Powerlessness of Necessity”, *Noûs*, 44(4), pp. 725–739. (Cited on pp. 222, 225.)

Schwenkler, John

2011 “Perception and Practical Knowledge”, *Philosophical Explorations*, 14(2), pp. 137–152. (Cited on pp. 170, 205.)

forthcoming “Understanding ‘Practical Knowledge’”, *Philosophers’ Imprint*. (Cited on pp. 178, 187, 205.)

Searle, John R.

1983 *Intentionality, an Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge. (Cited on pp. 104, 129, 193.)

1985 *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge. (Cited on p. 195.)

Searle, John R.

- 2001 *Rationality in Action*, MIT Press, Cambridge, Mass. (Cited on pp. 97, 195.)

Sehon, Scott R.

- 1997 “Deviant Causal Chains and the Irreducibility of Teleological Explanation”, *Pacific Philosophical Quarterly*, 78(2), pp. 195–213. (Cited on p. 123.)
- 2005 *Teleological Realism: Mind, Agency, and Explanation*, MIT Press, Cambridge, Mass. (Cited on pp. 140–142.)
- 2012 “Action Explanation and the Free Will Debate: How Incompatibilist Arguments Go Wrong”, *Philosophical Issues*, 22(1), pp. 351–368. (Cited on p. 142.)

Sellars, Wilfrid S.

- 1958 “Counterfactuals, Dispositions and the Causal Modalities”, in Feigl *et al.* [1958], vol. 2, pp. 225–308. (Cited on p. 216.)

Setiya, Kieran

- 2007 *Reasons without Rationalism*, Princeton University Press, Princeton. (Cited on p. 106.)
- 2008 “Practical Knowledge”, *Ethics*, 118(3), pp. 388–409. (Cited on p. 170.)
- 2011 “Knowledge of Intention”, in Ford *et al.* [2011], pp. 170–197. (Cited on p. 191.)
- 2012 “Knowing How”, *Proceedings of the Aristotelian Society*, 112(3), pp. 285–307. (Cited on p. 202.)

Shabo, Seth

- 2011 “Why Free Will Remains a Mystery”, *Pacific Philosophical Quarterly*, 92(1), pp. 105–125. (Cited on p. 69.)
- 2013 “Free Will and Mystery: Looking Past the Mind Argument”, *Philosophical Studies*, 162(2), pp. 291–307. (Cited on p. 69.)

Sider, Theodore

- 2001 *Four-Dimensionalism: An Ontology of Persistence and Time*, Oxford University Press, New York. (Cited on p. 228.)

Simons, Peter

- 1987 *Parts: A Study in Ontology*, Oxford University Press, New York. (Cited on p. 228.)

- Sinnott-Armstrong, Walter
- 2014 *Moral Psychology: Free Will and Moral Responsibility*, MIT Press, Cambridge, Mass. (Cited on p. 275.)
- Sinnott-Armstrong, Walter and Lynn Nadel
- 2010 (eds.), *Conscious Will and Responsibility: A Tribute to Benjamin Libet*, Oxford University Press, New York. (Cited on p. 280.)
- Slote, Michael
- 1982 “Selective Necessity and the Free-Will Problem”, *The Journal of Philosophy*, 79(1), pp. 5–24. (Cited on p. 78.)
- Smith, Kerri
- 2011 “Taking Aim at Free Will”, *Nature*, 477(7362), pp. 23–25. (Cited on pp. 50, 51.)
- Smith, Michael
- 1987 “The Humean Theory of Motivation”, *Mind*, 96(381), pp. 36–61. (Cited on p. 129.)
- 2004 “The Structure of Orthonomy”, *Royal Institute of Philosophy Supplement*, 55, pp. 165–193. (Cited on p. 94.)
- Soon, Chun Siong, Marcel Brass, Hans-Jochen Heinze and John-Dylan Haynes
- 2008 “Unconscious Determinants of Free Decisions in the Human Brain”, *Nature neuroscience*, 11(5), pp. 543–545. (Cited on pp. 14, 39, 40, 43–45.)
- Spence, Sean
- 1996 “Free Will in the Light of Neuropsychiatry”, *Philosophy, Psychiatry, & Psychology*, 3(2), pp. 75–90. (Cited on p. 27.)
- Steward, Helen
- 2008 “Fresh Starts”, *Proceedings of the Aristotelian Society*, 108(1), pp. 197–217. (Cited on p. 148.)
- 2012a *A Metaphysics for Freedom*, Oxford University Press, Oxford. (Cited on pp. 7, 10, 87, 99, 149–151, 153–155, 187.)
- 2012b “Actions as Processes”, *Philosophical Perspectives*, 26(1), pp. 373–388. (Cited on pp. 201, 226–228, 233.)
- Stoecker, Ralf
- 2003 “Climbers, Pigs and Wiggled Ears”, in Walter *et al.* [2003], pp. 296–322. (Cited on pp. 119, 120.)

Stout, Rowland

- 2010 “Deviant Causal Chains”, in O’Connor *et al.* [2010], pp. 159–165. (Cited on pp. 120, 121, 124, 125.)

Stoutland, Frederick

- 1989 “Von Wrights Theory of Action”, in Schilpp *et al.* [1989]. (Cited on p. 250.)

Strawson, Peter F.

- 1959 *Individuals*, Routledge, New York. (Cited on p. 6.)
1962 “Freedom and Resentment”, *Proceedings of the British Academy*, 48, pp. 1–25. (Cited on pp. 4, 5, 148.)

Stroud, Barry

- 2000 *The Quest for Reality: Subjectivism & the Metaphysics of Colour*, Oxford University Press, New York. (Cited on p. 4.)

Suppes, Patrick

- 1973 (ed.), *Logic, Methodology and Philosophy of Science IV*, Elsevier, Amsterdam. (Cited on p. 286.)

Swaab, Dick

- 2010 *Wij Zijn Ons Brein*, Atlas Contact, Amsterdam. (Cited on pp. 20, 23.)

Swinburne, Richard

- 2011 (ed.), *Free Will and Modern Science*, Oxford University Press, Oxford. (Cited on p. 261.)

Tanney, Julia

- 1995 “Why Reasons May Not be Causes”, *Mind & language*, 10(1-2), pp. 105–128. (Cited on p. 92.)
2005 “Reason-explanation and the Contents of the Mind”, *Ratio (new series)*, 18(3), pp. 338–351. (Cited on pp. 135, 137, 139.)
2009 “Reasons as Non-causal, Context-Placing Explanations”, in Sandis [2009]. (Cited on pp. 135, 136.)

Taylor, Richard

- 1973 *Action and Purpose*, Humanities Press, New York. (Cited on p. 149.)

Teichmann, Roger

- 2008 *The Philosophy of Elizabeth Anscombe*, Oxford University Press, Oxford. (Cited on pp. 170, 177, 188.)

Thalberg, Irving

- 1984 “Do Our Intentions Cause Our Intentional Actions?”, *American Philosophical Quarterly*, 21(3), pp. 249–260. (Cited on pp. 106, 108.)

Thompson, Michael

- 2008 *Life and Action*, Harvard University Press, Cambridge, Mass. (Cited on pp. 170, 174, 175, 200, 236.)
- 2011 “Anscombe’s *Intention* and Practical Knowledge”, in Ford *et al.* [2011], pp. 198–210. (Cited on pp. 180, 192, 201–203, 226.)

Tönnsjö, Torbjörn

- 2009 “On Deviant Causal Chains: No Need for a General Criterion”, *Analysis*, 69(3), pp. 469–473. (Cited on p. 126.)

Tooley, Michael

- 1977 “The Nature of Laws”, *Canadian Journal of Philosophy*, 7(4), pp. 667–98. (Cited on p. 149.)
- 1987 *Causation: A Realist Approach*, Clarendon Press, Oxford. (Cited on p. 149.)

Trevena, Judy and Jeff Miller

- 2010 “Brain Preparation Before a Voluntary Action: Evidence Against Unconscious Movement Initiation”, *Consciousness and Cognition*, 19(1), pp. 447–456. (Cited on pp. 29, 37, 40.)

Ulrich, Rolf, Judith Nitschke and Thomas Rammsayer

- 2006 “Perceived Duration of Expected and Unexpected Stimuli”, *Psychological research*, 70(2), pp. 77–87. (Cited on p. 26.)

van Inwagen, Peter

- 1986 *An Essay on Free Will*, Oxford University Press, Oxford. (Cited on pp. 66, 67, 78, 89, 148.)
- 2000 “Free Will Remains a Mystery: the Eighth Philosophical Perspectives Lecture”, *Philosophical Perspectives*, 14, pp. 1–19. (Cited on pp. 69, 73, 75.)

van Miltenburg, Niels

- 2012 “Practical Knowledge and Foreseen Side Effects”, *Journal of Ethics and Social Philosophy*. (Cited on p. 169.)

Velleman, J. David

- 1989 *Practical Reflection*, Princeton University Press, Princeton. (Cited on pp. 97, 179, 191.)

Velleman, J. David

- 1992a “The Guise of the Good”, *Noûs*, 26(1), pp. 3–26. (Cited on p. 195.)
- 1992b “What Happens When Someone Acts?”, *Mind*, 101(403), pp. 461–481. (Cited on p. 98.)
- 2000 *The Possibility of Practical Reason*, Oxford University Press. (Cited on p. 106.)
- 2004 “Précis of the Possibility of Practical Reason”, *Philosophical Studies*, 121, pp. 225–238. (Cited on p. 188.)

Velmans, Max and Susan Schneider

- 2007 (eds.), *The Blackwell Companion to Consciousness*, John Wiley & Sons, New York. (Cited on p. 260.)

Vogler, Candace

- 2001 “Anscombe on Practical Inference”, in Millgram [2001], pp. 437–464. (Cited on pp. 170, 178.)

Von Wright, Georg Henrik

- 1973 “On the Logic and Epistemology of the Causal Relation”, in Suppes [1973], pp. 293–312. (Cited on p. 249.)
- 1974 *Causality and Determinism*, Columbia University Press, New York. (Cited on pp. 249, 250.)

Walter, Sven and Heinz-Dieter Heckmann

- 2003 (eds.), *Physicalism and Mental Causation: The Metaphysics of Mind and Action*, Imprint Academic, Exeter. (Cited on p. 283.)

Watson, Gary

- 1982 (ed.), *Free Will*, Oxford University Press, Oxford. (Cited on p. 207.)

Wedgwood, Ralph

- 2006 “The Normative Force of Reasoning”, *Noûs*, 40(4), pp. 660–686. (Cited on pp. 113, 114.)

Wegner, Daniel M.

- 2002 *The Illusion of Conscious Will*, MIT press, Cambridge, Mass. (Cited on pp. 20, 24, 57.)
- 2008 “Self is Magic”, in Baer *et al.* [2008], pp. 226–247. (Cited on p. 57.)

Whittle, Ann

- 2008 “A Functionalist Theory of Properties”, *Philosophy and Phenomenological Research*, 77(1), pp. 59–82. (Cited on p. 220.)

Wiggins, David

- 1973 *Towards a Reasonable Libertarianism*, Chapter 8 in [Wiggins 1987]. (Cited on pp. 1, 13, 89, 247.)
- 1987 *Needs, Values, Truth: Essays in the Philosophy of Value*, Basil Blackwell, Oxford. (Cited on p. 287.)

Wilson, George

- 1985 “Davidson on Intentional Action”, in LePore *et al.* [1985], pp. 29–43. (Cited on p. 106.)

Wittgenstein, Ludwig

- 1953 *Philosophical Investigations*, Blackwell Publishing, Oxford. (Cited on pp. 133, 169.)

Woodward, James

- 2003 *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press, New York. (Cited on p. 249.)

Zhu, Jing

- 2003 “Reclaiming Volition: An Alternative Interpretation of Libet’s Experiment”, *Journal of Consciousness Studies*, 10(11), pp. 61–77. (Cited on p. 34.)

Samenvatting in het Nederlands

Wij zijn vrij. Mensen hebben het vermogen om zelf te kiezen welke handelingen zij uitvoeren. Filosofen noemen dit vermogen ‘vrije wil’. Maar ze zijn het niet eens over hoe we dit vermogen om vrij te handelen moeten begrijpen. In dit proefschrift tracht ik een zo intuïtief mogelijk plaatje van vrije wil te verdedigen: vrij handelen is het actualiseren van een open mogelijkheid. Voor mij staat nu een kopje koffie, daarom is het voor mij mogelijk om een slokje koffie te nemen. Als ik die slok neem verander ik een mogelijkheid in een realiteit. Stond de koffie er niet, dan had ik ook de mogelijkheid niet gehad om nu een slok te nemen (maar wel de mogelijkheid om koffie te gaan zetten en dus straks een slok te nemen). Door te handelen actualiseren we niet alleen mogelijkheden, we sluiten ook andere mogelijkheden af. Als ik mijn kopje koffie bijvoorbeeld leeg drink, maak ik het onmogelijk om nog een slok te nemen.

Volgens dit intuïtieve plaatje is de wereld als een tuin met vertakkende paden. Vrije wil is het vermogen om op elke kruising te bepalen welk pad we bewandelen en welke paden we links (of rechts) laten liggen. Veel filosofen vinden dit beeld van de wereld en de menselijke vrije wil nogal naïef en beweren dat het intuïtieve plaatje aangepast of zelfs geheel verworpen moet worden. Is het namelijk niet zo dat we gedwongen worden om een bepaald pad

te nemen? En zijn er eigenlijk altijd wel meerdere paden toegankelijk wanneer we handelen? Ook van buiten de filosofie komt er sinds kort kritiek op dit intuïtieve plaatje van vrije wil. Sommige neurowetenschappers menen namelijk dat niet wij bepalen welke paden er bewandeld worden, maar dat onze hersenen dat voor ons doen. Dit proefschrift legt uit hoe we het intuïtieve plaatje van vrije wil zo kunnen begrijpen dan het noch op filosofische noch op wetenschappelijke gronden verworpen of aangepast hoeft te worden.

In **Hoofdstuk 1** bespreek ik de meest recente neurowetenschappelijke experimenten waarvan beweerd wordt dat ze het bestaan van de vrije wil ontkrachten. In dit hoofdstuk bekritiseer ik deze experimenten en laat ik zien dat er een probleem is met de experimentele opzet dat er voor zorgt dat het principieel onmogelijk is om op basis van deze experimenten de vrije wil te verwerpen. De centrale reden waarom veel neurowetenschappers denken dat de vrije wil niet bestaat is dat ze in onze hersenen activiteit vinden die samenhangt met de keuzes die we gaan maken voordat we zo'n keuze daadwerkelijk maken. Aangezien we ons niet bewust zijn van deze vroege hersenactiviteit menen neurowetenschappers dat het een illusie is om te denken dat we keuzes maken: we worden ons slechts bewust van de keuzes die onze hersenen al lang gemaakt hebben. Het probleem met deze redenering is dat ze veronderstelt dat er niets mag gebeuren wat samenhangt met onze keuze voordat we deze gemaakt hebben. Maar dit is een verkeerde veronderstelling en we hebben geen hersenwetenschapper nodig om dit in te zien. Neem mijn beslissing om met mijn collega te gaan lunchen bijvoorbeeld. Deze beslissing hangt sterk samen met het feit dat hij net binnen kwam en me vroeg of ik mee ging lunchen. Dat maakt mijn beslissing om mee te gaan echter niet minder vrij: ik had ook kunnen weigeren omdat ik nog hard aan dit proefschrift moet werken. Voor de vrije wil is het dus niet zozeer van belang

dat wij ons bewust zijn van de eerste gebeurtenissen die samenhangen met ons handelen, het is van belang dat we zelf kunnen bepalen welke handeling we uit voeren. Ons handelen moet aan ons zijn, of met andere woorden: we moeten *controle* hebben.

Hoofdstuk 2 gaat over hoe verschillende filosofische posities controle begrijpen. Het hete hangijzer in het filosofische debat rondom vrije wil is niet zozeer of de vrije wil wel bestaat maar gaat over de vraag of de vrije wil indeterminisme verondersteld. Indeterminisme is het idee dat de toekomst open is—dat er meerdere verschillende dingen kunnen gebeuren. Determinisme is het idee dat alles al vanaf het begin ter tijden vast ligt. Om in termen van de tuinpaden te spreken: als determinisme waar is dan is er maar één pad, is indeterminisme waar dan zijn er splitsingen. In het intuïtieve plaatje van vrije wil zijn juist die splitsingen erg belangrijk, het intuïtieve plaatje is dus een indeterministisch plaatje. Er zijn ook filosofen (en die zijn in de meerderheid) die denken dat indeterminisme helemaal niet nodig is maar dat de vrije wil compatibel is met determinisme. Deze compatibilisten denken dat het eigenlijk geen enkel probleem is dat er maar een pad naar te toekomst leidt, zolang we maar op dat pad willen lopen en zolang we maar niet gedwongen worden om er te lopen, zijn we vrij.

Hoewel ik het intuïtieve plaatje van vrije wil verdedig en daarmee aan de kant van de indeterministen sta, denk ik dat de indeterministen in het huidige debat omtrent vrije wil een denkfout maken. Het probleem is dat deze de indeterministen nadenken over indeterminisme als iets wat onze controle over onze handelingen vergroot. Ze denken dat we best kunnen *handelen* als de wereld deterministisch zou zijn, maar dat we alleen niet *vrij* kunnen handelen. In dit hoofdstuk beargumenteer ik dat deze positie niet coherent is en dat indeterministen veel te ver mee gaan met een, zo ik beweer, inherent compatibilistische theorie van handelingen: de causale handelingstheorie. Als we het idee

dat vrije wil indeterminisme vereist serieus willen nemen moeten we niet denken dat indeterminisme controle vermeerdert of dat determinisme onze controle vermindert: we moeten denken dat determinisme onze controle vernietigt. Met andere woorden, het is niet alleen zo dat we indeterminisme nodig hebben om vrij te handelen, we hebben het nodig om überhaupt te kunnen handelen.

Het is daarom belangrijk dat de filosofen die het vrije wil debat voeren niet slechts over vrijheid nadenken, ze moeten vooral over handelingen nadenken. Dit gebeurt nu veel te weinig en dat is, beweer ik, de oorzaak van veel conceptuele verwarring.

Daarom gaat **Hoofdstuk 3** over handelingstheorie. Als iemand intentioneel handelt verandert er iets in de wereld: een arm gaat omhoog en er wordt een slok koffie genomen. Tegelijkertijd gebeuren intentionele handelingen omwille van een reden: ik neem een slok omdat ik bijvoorbeeld wakker wil blijven. Handelingen hebben dus zowel een plek in de fysieke wereld van gebeurtenissen als een plek in de mentale wereld van redenen. Maar hoe kan iets tegelijkertijd fysiek als rationeel zijn? Dit is wat handelingstheorie probeert te begrijpen.

Onder filosofen is het meest populaire verhaal over handelingen de bovengenoemde causale handelingstheorie. Deze theorie stelt dat handelingen rationeel kunnen zijn doordat ze kunnen overeenstemmen met de mentale toestanden van degene die handelt. Een slokje koffie nemen is bijvoorbeeld rationeel voor iemand die een verlangen heeft om wakker te blijven. Tegelijkertijd corresponderen mentale toestanden met fysieke toestanden in het hoofd. Een verlangen hebben is in een bepaalde lichamelijke toestand zijn. Volgens de causale handelingstheoreticus zijn het deze fysieke toestanden die onze handelingen veroorzaken en dat verklaart waarom handelingen fysieke gebeurtenissen zijn. Een intentionele handeling is dus een gebeurtenis die veroorzaakt wordt door een mentale toestand die een reden geeft voor de gebeurtenis.

Zoals ik in Hoofdstuk 2 beweer is deze causale theorie inherent compatibilistisch en daarom kan ze niet geaccepteerd worden door iemand die het intuïtieve plaatje van vrijheid verdedigen wil. De populariteit van de causale theorie is dus een slecht teken voor het indeterministische intuïtieve plaatje. Gelukkig voor dit plaatje is er iets fundamenteel mis met de causale handelingstheorie. Er zijn namelijk gebeurtenissen die zowel overeenstemmen met de reden van een persoon als veroorzaakt worden door deze reden, maar waarvan we toch niet willen zeggen dat het handelingen zijn. Stel je een sluipschutter in een bos voor die zijn doelwit wil vermoorden. Hij lost een schot maar mist zijn doel. Door de knal echter, wordt een nabije rotte wilde zwijnen zo opgeschrikt dat ze op hol slaan en toevallig het doelwit vertrappen met de dood tot gevolg. In dit verhaaltje wordt de dood van het doelwit veroorzaakt door een mentale toestand die een reden geeft voor die dood. (Het verlangen om het doelwit te doden veroorzaakt het overhalen van de trekker, wat een knal veroorzaakt, welke veroorzaakt dat de zwijnen op hol slaan, wat leidt tot de dodelijke vertrapping). Toch willen we niet zeggen dat het vertrappen van de zwijnen een intentionele handeling is van de sluipschutter.

Filosofen hebben zich in veel bochten gewrongen om te proberen het causale verhaal over handelingen te redden. Ze stellen dat een handeling niet slechts door een mentale toestand veroorzaakt moet worden maar dat dit ook ‘op de juiste wijze’ moet gebeuren. Maar wat is die juiste wijze? In dit hoofdstuk laat ik zien dat geen van de manieren waarop filosofen dit proberen uit te leggen echt werkt. Dit falen is geen toeval maar een symptoom van een dieper probleem voor de causale handelingstheorie. Het probleem is dat deze handelingstheorie het rationele en het fysieke van een handeling los van elkaar probeert te verklaren: het rationele in termen van overeenstemming tussen de handeling en de inhoud van een mentale toestand, en het fysieke gebeuren in termen van ver-

oorzaking. Dit terwijl deze twee aspecten bij een handeling juist niet los staan: een handeling wordt gedaan omdat daar reden voor is. De conclusie is dat de causale handelingstheorie moet worden verworpen.

In **Hoofdstuk 4** beschouw ik een aantal recent geopperde alternatieven voor de causale handelingstheorie en tracht ik aan te tonen dat geen van deze alternatieven met een bevredigend verhaal komt. Alternatieven voor de causale theorie zijn er in twee smaken: de niet-causale theorie, en een theorie die stelt dat handelingen niet door de mentale toestanden van een handelende persoon, maar direct door de persoon zelf veroorzaakt worden. De niet-causale theorie werkt niet, precies omdat ze niet causaal is, en daardoor geen goed verhaal kan hebben over waarom handelingen fysieke gebeurtenissen zijn. Het probleem met de huidige theorieën van directe veroorzaking door de handelende persoon zelf, is dat ze niet goed kunnen uitleggen hoe die veroorzaking door de persoon samenhangt met de redenen van de persoon: het rationele aspect van handelen blijft dus onvoldoende belicht.

Hoe moeten we handelen dan wel begrijpen? In **Hoofdstuk 5** verdedig ik een handelingstheorie in termen van praktische kennis. Het verschil tussen intentionele handeling en andere lichaamsbewegingen is dat je van intentionele handelingen kennis hebt op het moment dat je ze uitvoert. Als je in een volle bus door de bocht gaat en je gewicht verplaatst kan het voorkomen dat je per ongeluk op iemands teen gaat staan. Dit is precies per ongeluk omdat je niet wist dat je op die teen zou gaan staan door je gewicht te verplaatsen. Ging je expres op die teen staan, dan wist je heel goed waar je mee bezig was. Volgens de filosofe Elizabeth Anscombe is deze praktische kennis die je hebt van je intentionele handelingen essentieel voor deze handelingen: ze maken de handelingen tot wat ze zijn. Neem Sacha die uien aan het snijden is in de keuken. Van buiten af kunnen we nu niet zien of hij bij-

voorbeeld spaghetti maakt of risotto. Maar voor Sacha zelf is dit allang duidelijk, hij weet dat hij uien snijdt omdat hij risotto wil maken. De kennis die Sacha heeft op het moment dat hij snijdt maakt zijn snijden tot onderdeel van het risotto koken. Daarbij snijdt hij alleen uien omdat hij weet dat hij risotto aan het koken is; zou hij een appeltaart aan het bakken zijn dan zou hij geen uien snijden. De handeling (het snijden van uien) gebeurt omdat ze rationeel is (Sacha weet dat de handeling een middel is tot zijn doel: een overheerlijke risotto).

Veel filosofen hebben echter een probleem met deze theorie. Zij denken dat het onzin is om te stellen dat handelen door kennis gekarakteriseerd wordt omdat we soms helemaal niet weten wat we aan het doen zijn. Iemand kan denken dat hij een notitie maakt zonder op zijn kladblok te kijken, maar als zonder dat hij dit weet de inkt in zijn pen op raakt, dan maakt hij maar een halve notitie en weet hij dus zelf niet wat hij doet. Toch is het zo dat hij intentioneel met zijn (lege) pen over zijn papier beweegt. Volgens mij geeft dit echter een verkeerde voorstelling van zaken. Handelingen zijn geen statische gebeurtenissen die er in één keer helemaal zijn. Het zijn dynamische processen die bestaan uit verschillende fasen. Een proces kan aan de gang zijn zonder dat het al voltooid is of zelfs zonder dat het ooit tot voltooiing komt. De bal is naar de grond aan het vallen, maar ik vang hem op voordat hij de grond bereikt en dus is de bal uiteindelijk niet naar de grond gevallen. Het geval van de notulist met de lege pen lijkt hier op, de notulist is aan het schrijven maar momenteel wordt er even niks geschreven omdat de pen op is. Dat hij wel aan het schrijven is blijkt wel uit het feit dat hij, zodra hij door heeft dat zijn pen leeg is, de vulling verwisselt en de notie afmaakt. Het opmerken dat de pen leeg is en het verwisselen van de vulling zijn simpelweg onderdeel van het overkoepelende schrijfproces. Het is dus geen onzin om te denken dat mensen altijd kennis van hun intentionele handelingen

hebben zolang we deze kennis maar opvatten als kennis van een proces wat niet al af is, maar wat nog aan het gebeuren is. We weten altijd wat we aan het doen zijn maar vaak niet of we het ook gedaan krijgen.

Hoewel praktische kennis uitlegt hoe een handeling zowel rationeel als materieel kan zijn, blijft het bijzonder dat wij mensen in staat zijn om deze bewegingen te produceren. In **Hoofdstuk 6** leg ik uit dat we ons vermogen om te handelen het beste kunnen begrijpen binnen een bredere theorie van vermogens. Voor veel filosofen zijn vermogens problematisch. Neem bijvoorbeeld het vermogen van een wijnglas om te breken. Breekbaarheid is net zoals transparantie een van de eigenschappen van het wijnglas, maar in tegenstelling tot transparantie kun je niet direct zien dat een glas deze eigenschap heeft. De breekbaarheid komt pas tot uiting als het glas bijvoorbeeld op de grond valt en daadwerkelijk breekt. Soms komt een vermogen zelfs nooit tot uiting, bijvoorbeeld als het glas omgesmolten wordt voordat het ooit breekt. We kunnen zeggen dat breekbaarheid dus een eigenschap is die een potentie aangeeft terwijl transparantie een eigenschap is die altijd actueel is. Veel filosofen hebben het idee dat vermogens op een of andere manier ‘minder echt’ zijn dan altijd actuele eigenschappen en proberen vermogens te reduceren tot zulke echtere eigenschappen. Ik geef argumenten tegen een dergelijke reductie en probeer een plausibele theorie van vermogens te geven die stelt dat vermogens net zo echt zijn als andere eigenschappen.

In de filosofische literatuur bestaat er discussie over een aantal technische problemen voor theorieën van vermogens. Het blijkt erg moeilijk om uit te leggen hoe het kan dat vermogens soms niet tot uiting komen zelfs als daarvoor de juiste condities aanwezig zijn. Een glas kan breken als het valt, maar dit hoeft niet. Ik geef een nieuwe oplossing voor dit probleem. Mijn oplossing rust op het idee dat ook in het debat omtrent vermogens teveel over

gebeurtenissen wordt nagedacht als statische entiteiten in plaats van als dynamische processen.

Een theorie van vermogens past erg goed bij het intuïtieve plaatje van vrije wil. Ik kan een slok koffie nemen omdat het kopje dat voor mij staat het vermogen heeft om koffie te bevatten en omdat ik sterk genoeg ben om het kopje op te tillen. Zo bepalen onze eigen vermogens en de vermogens van de dingen om ons heen wat mogelijk is. Denk nu terug aan het beeld van de tuin met vertakkende paden. De paden representeren onze echte mogelijkheden. We kunnen nu inzien dat deze paden gevormd worden door onze vermogens en de vermogens van de dingen om ons heen. Welk pad we bewandelen hangt af van welke vermogens uitgeoefend worden. Door ons eigen vermogen tot handelen te gebruiken bepalen we dus wat er gebeurt. Op deze manier beïnvloeden we de loop van het universum en daarin ligt onze vrijheid besloten.

Curriculum Vitae

Niels van Miltenburg was born on June 22nd 1986 in Nieuwegein, the Netherlands. In 2009 he graduated at Utrecht University with a bachelors degree in philosophy and a minor in artificial intelligence. He was admitted to the research master program and obtained his master degree in philosophy in 2011 *cum laude*. From 2011 to 2013 he was a PhD researcher in Utrecht for Thomas Müller's research program *Indeterminism Ltd.* which was funded by the European Research Council. When the project moved to the university of Konstanz he was appointed by that university to continue his research. He has taught courses on topics in the philosophy of mind, metaphysics and the philosophy of action, and has given talks on free will in Amsterdam, Bonn, Cologne, Oslo and Oxford.

When Niels is not doing philosophy he enjoys the company of his friends, playing the trombone, games, watching sports and good books. He happily lives in Utrecht with his wife, Belén, and their cat, Poes.

Quaestiones Infinitae

PUBLICATIONS OF THE DEPARTMENT OF PHILOSOPHY AND RELIGIOUS STUDIES

- VOLUME 28. J.A. Bergstra, S.F.M. van Vlijmen, *Theoretische Software-Engineering: kenmerken, faseringen en classificaties*, 1998.
- VOLUME 29. A.G. Wouters, *Explanation Without A Cause* (dissertation), 1999.
- VOLUME 30. M.M.S.K. Sie, *Responsibility, Blameworthy Action & Normative Disagreements* (dissertation), 1999.
- VOLUME 31. M.S.P.R. van Atten, *Phenomenology of choice sequences* (dissertation), 1999.
- VOLUME 32. V.N. Stebletsova, *Algebras, Relations and Geometries (an equational perspective)* (dissertation), 2000.
- VOLUME 33. A. Visser, *Het Tekst Continuüm* (inaugural lecture), 2000.
- VOLUME 34. H. Ishiguro, *Can we speak about what cannot be said?* (public lecture), 2000.
- VOLUME 35. W. Haas, *Haltlosigkeit; Zwischen Sprache und Erfahrung* (dissertation), 2001.
- VOLUME 36. R. Poli, *ALWIS: Ontology for knowledge engineers* (dissertation), 2001.
- VOLUME 37. J. Mansfeld, *Platonische Briefschrijverij* (valedictory lecture), 2001.
- VOLUME 37a. E.J. Bos, *The Correspondence between Descartes and Henricus Regius* (dissertation), 2002.
- VOLUME 38. M. van Otegem, *A Bibliography of the Works of Descartes (1637-1704)* (dissertation), 2002.
- VOLUME 39. B.E.K.J. Goossens, *Edmund Husserl: Einleitung in die Philosophie: Vorlesungen 1922/23* (dissertation), 2003.
- VOLUME 40. H.J.M. Broekhuijse, *Het einde van de sociaaldemocratie* (dissertation), 2002.
- VOLUME 41. P. Ravalli, *Husserls Phänomenologie der Intersubjektivität in den Göttinger Jahren: Eine kritisch-historische Darstellung* (dissertation), 2003.
- VOLUME 42. B. Almond, *The Midas Touch: Ethics, Science and our Human Future* (inaugural lecture), 2003.
- VOLUME 43. M. Düwell, *Morele kennis: over de mogelijkheden van toegepaste ethiek* (inaugural lecture), 2003.
- VOLUME 44. R.D.A. Hendriks, *Metamathematics in Coq* (dissertation), 2003.
- VOLUME 45. Th. Verbeek, E.J. Bos, J.M.M. van de Ven, *The Correspondence of René Descartes: 1643*, 2003.
- VOLUME 46. J.J.C. Kuiper, *Ideas and Explorations: Brouwer's Road to Intuitionism* (dissertation), 2004.
- VOLUME 47. C.M. Bekker, *Rechtvaardigheid, Onpartijdigheid, Gender en Sociale Diversiteit; Feministische filosofen over recht doen aan vrouwen en hun onderlinge verschillen* (dissertation), 2004.

- VOLUME 48. A.A. Long, *Epictetus on understanding and managing emotions* (public lecture), 2004.
- VOLUME 49. J.J. Joosten, *Interpretability formalized* (dissertation), 2004.
- VOLUME 50. J.G. Sijmons, *Phänomenologie und Idealismus: Analyse der Struktur und Methode der Philosophie Rudolf Steiners* (dissertation), 2005.
- VOLUME 51. J.H. Hoogstad, *Time tracks* (dissertation), 2005.
- VOLUME 52. M.A. van den Hoven, *A Claim for Reasonable Morality* (dissertation), 2006.
- VOLUME 53. C. Vermeulen, *René Descartes, Specimina philosophiae: Introduction and Critical Edition* (dissertation), 2007.
- VOLUME 54. R.G. Millikan, *Learning Language without having a theory of mind* (inaugural lecture), 2007.
- VOLUME 55. R.J.G. Claassen, *The Market's Place in the Provision of Goods* (dissertation), 2008.
- VOLUME 56. H.J.S. Bruggink, *Equivalence of Reductions in Higher-Order Rewriting* (dissertation), 2008.
- VOLUME 57. A. Kalis, *Failures of agency* (dissertation), 2009.
- VOLUME 58. S. Graumann, *Assistierte Freiheit* (dissertation), 2009.
- VOLUME 59. M. Aalderink, *Philosophy, Scientific Knowledge, and Concept Formation in Geulincx and Descartes* (dissertation), 2010.
- VOLUME 60. I.M. Conradie, *Seneca in his cultural and literary context: Selected moral letters on the body* (dissertation), 2010.
- VOLUME 61. C. van Sijl, *Stoic Philosophy and the Exegesis of Myth* (dissertation), 2010.
- VOLUME 62. J.M.I.M. Leo, *The Logical Structure of Relations* (dissertation), 2010.
- VOLUME 63. M.S.A. van Houte, *Seneca's theology in its philosophical context* (dissertation), 2010.
- VOLUME 64. F.A. Bakker, *Three Studies in Epicurean Cosmology* (dissertation), 2010.
- VOLUME 65. T. Fossen, *Political legitimacy and the pragmatic turn* (dissertation), 2011.
- VOLUME 66. T. Visak, *Killing happy animals. Explorations in utilitarian ethics.* (dissertation), 2011.
- VOLUME 67. A. Joosse, *Why we need others: Platonic and Stoic models of friendship and self-understanding* (dissertation), 2011.
- VOLUME 68. N.M. Nijsingh, *Expanding newborn screening programmes and strengthening informed consent* (dissertation), 2012.
- VOLUME 69. R. Peels, *Believing Responsibly: Intellectual Obligations and Doxastic Excuses* (dissertation), 2012.
- VOLUME 70. S. Lutz, *Criteria of Empirical Significance* (dissertation), 2012.
- VOLUME 70a. G.H. Bos, *Agential Self-consciousness, beyond conscious agency* (dissertation), 2013.

- VOLUME 71. F.E. Kaldewaij, *The animal in morality: Justifying duties to animals in Kantian moral philosophy* (dissertation), 2013.
- VOLUME 72. R.O. Buning, *Henricus Reneri (1593-1639): Descartes' Quartermaster in Aristotelian Territory* (dissertation), 2013.
- VOLUME 73. I.S. Löwisch, *Genealogy Composition in Response to Trauma: Gender and Memory in 1 Chronicles 1–9 and the Documentary Film 'My Life Part 2'* (dissertation), 2013.
- VOLUME 74. A. El Khairat, *Contesting Boundaries: Satire in Contemporary Morocco* (dissertation), 2013.
- VOLUME 75. A. Krom, *Not to be sneezed at. On the possibility of justifying infectious disease control by appealing to a mid-level harm principle* (dissertation), 2014.
- VOLUME 76. Z. Pall, *Salafism in Lebanon: local and transnational resources* (dissertation), 2014.
- VOLUME 77. D. Wahid, *Nurturing the Salafi Manhaj: A Study of Salafi Pesantrens in Contemporary Indonesia* (dissertation), 2014.
- VOLUME 78. B.W.P. van den Berg, *Speelruimte voor dialoog en verbeelding. Basisschoolleerlingen maken kennis met religieuze verhalen* (dissertation), 2014.
- VOLUME 79. J.T. Berghuijs, *New Spirituality and Social Engagement* (dissertation), 2014.
- VOLUME 80. A. Wetter, *Judging By Her. Reconfiguring Israel in Ruth, Esther and Judith* (dissertation), 2014.
- VOLUME 81. J.M. Mulder, *Conceptual Realism. The Structure of Metaphysical Thought* (dissertation), 2014.
- VOLUME 82. L.W.C. van Lit, *Eschatology and the World of Image in Suhrawardī and His Commentators* (dissertation), 2014.
- VOLUME 83. P.L. Lambertz, *Divisive matters Aesthetic difference and authority production in a Congolese spiritual movement "from Japan"* (dissertation), 2015 .
- VOLUME 84. J.P. Goudsmit, *Intuitionistic Rules: Admissible Rules of Intermediate Logics* (dissertation), 2015.
- VOLUME 85. E.T. Feikema, *Still not at Ease: Corruption and Conflict of Interest in Hybrid Political Orders* (dissertation), 2015.
- VOLUME 86. N. van Miltenburg, *Freedom in Action* (dissertation), 2015.