



OXFORD STUDIES IN

Agency and
Responsibility

Volume 3

OXFORD

OXFORD STUDIES IN AGENCY AND
RESPONSIBILITY

Oxford Studies in Agency and Responsibility

Volume 3

Edited by
DAVID SHOEMAKER

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

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

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

© The several contributors 2015

The moral rights of the authors have been asserted

First Edition published in 2015

Impression: 1

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

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

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

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2015937038

ISBN 978-0-19-874483-2 (hbk.)

ISBN 978-0-19-874484-9 (pbk.)

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

Links to third party websites are provided by Oxford in good faith and
for information only. Oxford disclaims any responsibility for the materials
contained in any third party website referenced in this work.

Acknowledgments

Versions of the articles in this collection were presented at the second biennial New Orleans Workshop on Agency and Responsibility (NOWAR), November 7–9, 2013. I would like to thank the Murphy Institute at Tulane University, and its director Steven Sheffrin, for generous financial support of this ongoing endeavor. Thanks to Meg Keenan for logistical support, and to Nate Stout and Frankie Worrell for their help with the execution of the workshop. Thanks to the panel of referees who helped me vet the workshop abstracts: David Brink, Luca Ferrero, Dana Nelkin, Paul Russell, Tamler Sommers, and Matt Talbert. Thanks to two anonymous referees at Oxford University Press, who provided valuable feedback on the initial drafts of the papers. Thanks to Nate Stout, who compiled the index (his third for the OSAR series). Thanks to the team at OUP—copy editors, publicists, proofreaders, managers—who all contributed something positive to the volume. And thanks, finally, to Peter Momtchiloff, my Oxford editor, for his continual support of this series.

David Shoemaker
New Orleans, LA
May 2015

Contents

<i>List of Contributors</i>	ix
Introduction <i>David Shoemaker</i>	1
1 Free Will and Agential Powers <i>Randolph Clarke and Thomas Reed</i>	6
2 Injecting the Phenomenology of Agency into the Free Will Debate <i>Terry Horgan</i>	34
3 Coherence of Attitudes, Integration of the Self, and Personal Integrity <i>Sigrún Svavarsdóttir</i>	62
4 Obsessive-Compulsive Disorder and Moral Agency <i>Timothy Schroeder</i>	85
5 Sensitivity to Reasons and Actual Sequences <i>Carolina Sartorio</i>	104
6 Responsibility and the Actual Sequence <i>John Martin Fischer</i>	120
7 Moral Luck Reexamined <i>Michael J. Zimmerman</i>	136
8 The Hard Problem of Responsibility <i>Victoria McGeer and Philip Pettit</i>	160
9 Rationality, Authority, and Bindingness: An Account of Communal Norms <i>Bennett W. Helm</i>	189
10 A Difference-Making Framework for Intuitive Judgments of Responsibility <i>David Lagnado and Tobias Gerstenberg</i>	213
11 Moral Responsibility, Reasons, and the Self <i>Chandra Sripada</i>	242
<i>Index</i>	265

List of Contributors

Randolph Clarke is Professor of Philosophy at Florida State University.

John Martin Fischer is Distinguished Professor of Philosophy at University of California, Riverside.

Tobias Gerstenberg is a Postdoctoral Fellow in Brain and Cognitive Sciences at MIT.

Bennett W. Helm is the Elijah E. Kresge Professor of Philosophy at Franklin & Marshall College.

Terry Horgan is Professor of Philosophy at University of Arizona.

David Lagnado is Senior Lecturer in the Division of Psychology and Language Sciences at University College London.

Victoria McGeer is Research Scholar at the University Center for Human Values and Lecturer in the Department of Philosophy of Princeton University, and is also in the School of Philosophy at Australian National University.

Philip Pettit is Laurance S. Rockefeller Professor of Politics and Human Values at Princeton University, and is also in the School of Philosophy at Australian National University.

Thomas Reed is a graduate student at Florida State University.

Carolina Sartorio is Associate Professor of Philosophy at University of Arizona.

Timothy Schroeder is Professor of Philosophy at Ohio State University.

Chandra Sripada is Assistant Professor of Philosophy and Assistant Professor of Psychiatry at University of Michigan.

Sigrún Svavarsdóttir is Associate Professor of Philosophy at Tufts University.

Michael J. Zimmerman is Professor of Philosophy at University of North Carolina, Greensboro.

Introduction

David Shoemaker

This is the third volume of *Oxford Studies in Agency and Responsibility* (OSAR), and the second drawn from papers presented at the New Orleans Workshop on Agency and Responsibility (NOWAR). The OSAR series is devoted to publishing cutting-edge, interdisciplinary work on the wide array of titular topics. In this volume, there are papers on free agency, agential integrity, the phenomenology of agency, reasons-responsiveness, moral luck, moral scrupulosity, accountability to community, responsibility and the deep self, and the psychology of responsibility judgments. In what follows, I briefly introduce the essays and explain their place in the dialectic.

The first half of the volume is made up of essays that fall roughly under the rubric of agency, and the second half is made up of essays that fall roughly under the rubric of responsibility. In “Free Will and Agential Powers,” Randolph Clarke and Thomas Reed aim to offer and assess the prospect of understanding free will in terms of the *powers* of agents. Powers may be many different kinds of things, but the option Clarke and Reed explore is powers as causal dispositions. They point out some hurdles such an account must overcome, such as capturing the crucial “up to us” component of free will, and they point out some possible avenues for doing so. But their primary goal is simply to put on the table a robust and clear-eyed view of the agential powers construal of free will, one on an equal footing for assessment with others.

Many free will theorists have taken the phenomenology of agency—what it is like from the inside to be an agent—to favor libertarianism, the view that the type of free will agents like us in fact enjoy is incompatible with determinism (see, e.g., Campbell 1951; Deery, Bedke, and Nichols 2013). In “Injecting the Phenomenology of Agency into the Free Will Debate,” Terry Horgan attempts to turn this view on its head. He argues for including phenomenological experience in our theorizing about free will, and in particular he draws our attention to two phenomenological features any

plausible theory must incorporate, namely, our experience of various bodily motions as our *own* movements, and our experience of actions as things we could have done otherwise. Horgan then shows how an adequately robust position on free will must account for how this phenomenological dimension interacts with several other dimensions of the debate, and that the best package deal theoretical account of this interaction is not libertarianism, but compatibilism.

Sigrún Svavarsdóttir, in “Coherence of Attitudes, Integration of the Self, and Personal Integrity,” focuses on explicating a particular type of agent, namely, the agent with integrity. She argues against a standard view of personal integrity that essentially conflates it with integration of the self, such that both are about *oneness*, wherein various attitudes are integrated into a coherent whole. While this is indeed true of integration of the self, she argues, personal integrity is about standing up for one’s commitments, perhaps despite one’s being poorly integrated, and that doing so is a function of our special capacity for reflection on those commitments.

In “Obsessive-Compulsive Disorder and Moral Agency,” Timothy Schroeder explores a fascinating case study in pathological agency. Agents with *moral scrupulosity* are obsessed with adhering to moral norms, as well as atoning for what they take to be their failures to do so, and an important feature of the disorder is that they take these obsessions to be reflections of who they are and what they want, rather than as alien forces (which is how many other agents with Obsessive-Compulsive Disorder (OCD) view their obsessions). What they intrinsically desire, it seems, is to do the right thing. But surely, we would tend to think, they are not morally responsible, that is, not apt subjects of praise or blame. These two data points, however, seem to leave neo-Humean theories of practical reasons in dire straits, for on such theories, one is morally responsible only if one acts for a reason, but to act for a reason is to act on a desire. If the morally scrupulous do act on desires, therefore, they would be eligible for moral responsibility on this theory, which seems the wrong answer. Schroeder explores whether the neo-Humean may have the resources to deny that the obsessions of those with scrupulosity are truly desires by paying careful attention to the neuropsychological literature as well as phenomenological reports. There is some possibility that the neural basis of OCD (and so of the morally scrupulous) is distinct from what one would expect if it were a disorder with unusually strong intrinsic desires, for instance, although Schroeder’s refreshingly honest conclusion is that it may just be too soon to tell one way or the other how, or even if, the neo-Humean will have a plausible response to this worry.

In “Sensitivity to Reasons and Actual Sequences,” Carolina Sartorio focuses on two conditions of free agency that seem to be in tension: (a) acting freely requires sensitivity to reasons, and (b) acting freely is exclusively

a function of the agent's actual causal history. The tension between these requirements arises from the fact that (a) is modal, referring to the agent's sensitivity to reasons in other possible worlds, and (b) isn't, referring to the reasons the agent actually had. Sartorio attempts to resolve this tension by appealing to the *absence* of reasons and how they may play a causal role in the actual sequence.

While Sartorio focuses on freedom in the actual sequence of events, John Martin Fischer focuses on *responsibility* in the actual sequence. Fischer has long been a leading advocate of the view that moral responsibility does not require the freedom to do otherwise (what he calls regulative control); all it requires instead is the more compatibilist-friendly condition that Sartorio puts forward as her second requirement above, namely, it must be a function of the actual sequence of what led to the agent's behavior, where for Fischer this actual sequence could have a modal character (being, say, a function of a reasons-responsive mechanism capable of responding differently to alternative reasons). Fischer's contribution to this volume—"Responsibility and the Actual Sequence"—addresses a significant challenge to his "Semicompatibilist" view, an objection that the Frankfurt-style cases on which Fischer relies feature counterfactual interveners, whereas causal determination is a property of the way the actual sequence unfolds. Fischer's carefully calibrated response involves articulating just how a factor that might render someone unable to do otherwise could still be irrelevant in accounting for his action, so that it would be "gratuitous," in Harry Frankfurt's phrase, "to assign it any weight in the assessment of his moral responsibility" (Frankfurt 1988: 8).

In "Moral Luck Reexamined," Michael Zimmerman also focuses on the relation between moral responsibility and control. Many people accept that luck about the results of our decisions—whether, say, I succeed in hitting the person I decided to shoot at rather than just a bird that flew in between us—cannot affect my moral responsibility, for these results are outside of our control, and what is outside of our control cannot bear one way or the other on our moral responsibility. But people generally embrace the thought that the luck about which *situations* we find ourselves in—our circumstances or even our characters—can bear on our moral responsibility. Zimmerman's radical view is that situational luck has exactly the same bearing on our moral responsibility as resultant luck, namely, none. I may bear the exact same culpability if I *would have* shot someone but for my timidity as I do if I actually successfully shot that person. In this essay, Zimmerman defends and develops this view in light of a sly and subtle challenge.

In "The Hard Problem of Responsibility," Victoria McGeer and Philip Pettit tackle a deeply puzzling issue for those who adhere to a reasons-responsive theory of responsibility, namely, how it can account for

blameworthy wrongdoing. On the standard version of the theory, one is responsible—fit to be held responsible—for an action to the extent one had the capacity to recognize and respond to reasons to do otherwise but failed to do so without excuse. The hard problem, McGeer and Pettit point out, is to figure out how such a failure could be the result of anything other than a glitch or an accident, in which case the agent would *not* be the apt object of the sort of emotional responses like resentment and indignation in which holding her responsible consists. In responding to this problem, the authors argue that when we ascribe a reasons-responsive capacity to a wrongdoing agent via emotional blame, we are actually *capacitating* that agent, that is, “supporting and reinforcing” the capacity we are ascribing to her. The hard problem may be resolved by seeing that there are actually two sensitivities at work in our being responsive to reasons, namely, a sensitivity to reasons, of course, but also a sensitivity to the *audience* of our actions, and their emotional responses to our failures may serve to render our disposition to respond to reasons more robust.

Bennett Helm, in “Rationality, Authority, and Bindingness: An Account of Communal Norms,” also emphasizes the importance of the audience—our fellow community members—in accountability. He argues that the bindingness of community norms, which includes the authority that our fellows have to hold us accountable for flouting them, has its source in our joint *commitment* to that community, as well as to each other. This commitment is implicated in the wide range of reactive attitudes we have to one another. These reactive attitudes are all part of a web of rationally interconnected attitudes, and when I take them up, I both assert my authority to demand compliance from others and recognize the authority of others to do the same to me. When the other members of the community take up this call as well, we have formed the joint commitment in which our binding ourselves and others to the communal norms consists.

In contrast with these normative explorations of responsibility, David Lagnado and Tobias Gerstenberg offer a theory of our actual responsibility judgments in “A Difference-Making Framework for Intuitive Judgments of Responsibility.” In particular, they are interested in explaining what is going on in people’s attributions of responsibility in cases of multiple agents producing joint outcomes. Drawing from several empirical studies, they theorize that people’s assignments of responsibility to individual agents in these scenarios are a function of both the extent to which the agents are viewed as critical for the outcome before the fact, and the extent to which they were thought to have made a difference to the outcome after the fact. That is, what seems to matter is the overall *difference* the agent was thought to have made, and Lagnado and Gerstenberg show that this understanding

of difference-making is based on a modified notion of counterfactual dependence.

There are, roughly, two traditions of conceptualizing and theorizing about moral responsibility. On the one hand, there are reasons-responsive views (represented in this volume by Sartorio, Fischer, and McGeer and Pettit), according to which an agent's responsibility is a matter of her being able to recognize and respond to the relevant moral reasons. On the other hand, there are deep self views (represented in Frankfurt 1988; Watson 2004; and Smith 2005), according to which an agent's responsibility is a matter of her actions and attitudes expressing her deep self, that subset of her psychic elements taken to represent who she really is, in some specified sense. These two traditions offer competing grounds for responsibility and non-responsibility, and they occasionally spit out competing verdicts. In "Moral Responsibility, Reasons, and the Self," Chandra Sripada attempts a rapprochement between the two approaches by suggesting that they may both be drawing on a Humean account of reasons. On Sripada's understanding of the Humean account, reasons for action are conceptually connected to the desires that make up the person's self. If this is right, then both traditions could be construed as saying that moral responsibility is really about expressing one's self through action, and this could bridge the gap between them.

References

- Campbell, C. A. (1951). "Is 'Freewill' a Pseudo-problem?" *Mind* 60(240): 441–65.
- Deery, Oisín, Bedke, Matt, and Nichols, Shaun. (2013). "Phenomenal Abilities: Incompatibilism and the Experience of Agency." *Oxford Studies in Agency and Responsibility* 1: 126–50.
- Frankfurt, Harry. (1988). *The Importance of What We Care About* (Cambridge: Cambridge University Press).
- Smith, Angela. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115: 236–71.
- Watson, Gary. (2004). *Agency and Answerability* (Oxford: Oxford University Press).

Free Will and Agential Powers

Randolph Clarke and Thomas Reed

What is free will, or (as it might be called) freedom of the will? It was once common to think of it as a power (or powers) of some kind. Locke, Hume, and Reid, for example, despite advancing very different accounts, all couched their views in terms of powers.¹

Fashions come and go. And come again. Lately several theorists have revived talk of powers or dispositions in their theories of free will.² There's good reason to think they're on to something. Even when not explicitly invoked, powers are often implicit in what accounts of free will require. Powers of some kind are implicated in an agent's having rational capacities—capacities to recognize and weigh reasons, to believe, desire, intend, and act for reasons, and to engage in practical reasoning and make decisions on its basis—which virtually every theorist takes to be required. The reasons-responsiveness of an action-producing mechanism (which is the central requirement of one influential contemporary view)³ is that mechanism's instantiation of a certain power. Being able to perform an action of one type or another is commonly advanced as a requirement; and having such an ability is at least partly a matter of having certain powers. Even a characterization of free will in terms of the *absence* of obstacles or impediments implicitly appeals to powers, for an obstacle limits one's freedom only to the extent that one lacks a power to circumvent or surmount it.⁴

Several recent accounts that make explicit appeal to powers are beset by problems of one sort or another.⁵ Rather than engage in critique, here we'll

¹ See Locke, *Essay*, Bk. 2, Ch. 21; Hume, *Enquiry*, Sect. 8; and Reid, *Essays*, Essay IV.

² Some examples are Fara (2008); Goetz (2008); Lowe (2008); Pink (2004, 2011); and Vihvelin (2004, 2013: ch. 6).

³ The view is that of Fischer and Ravizza (1998).

⁴ Van Inwagen (2008b: 460) makes a similar point in terms of ability rather than power.

⁵ For critical examination of these views, see Clarke (2009) and Whittle (2010).

offer some proposals of our own. Our aim is to contribute to a balanced assessment of the prospects for one very attractive approach to free will.

We'll begin by offering a construal of *willing to do a certain thing*, with *willing* understood to be an action with respect to which an agent might be free. We'll then identify what we take to be a constraint on any theory of free will that aims to capture our ordinary notion of the phenomenon. We'll discuss conceptions of powers that might be employed in an account of free will. Many such views invoke powers that are in an important respect like the dispositions of inanimate objects. Following this lead, we'll offer several suggestions concerning how an account of this type might go, and we'll discuss a problem that any such view faces. We'll conclude with brief consideration of alternative approaches that invoke powers of quite different kinds.⁶

1.1 WILLING

Locke's own treatment of what he called liberty or freedom illustrates a typical shortcoming of accounts expressed in terms of powers. His initial—and his main—proposal characterized freedom with respect to *doing what one might will to do*, not freedom with respect to *willing to do a certain thing*. One is at liberty to *A*, Locke said, just in case one has a power to *A* or not *A*, according to which of these one wills to do. And as he pointed out, being at liberty to *will to A* can't sensibly be understood in the same fashion.⁷

Now, our exercise of free will might well be, in part, our freedom in doing what we will to do. But we submit that free will is, in the first instance, freedom with respect to willing to do this or that.⁸ Shortly we'll advance an understanding of willing appropriate for this view of free will.

⁶ We take no stand here on whether powers are basic features of the world or, instead, reducible in one way or another. Our main proposals are consistent with various views on this question.

⁷ Several writers who advance conceptions of abilities to act in terms of powers likewise advance analyses of abilities to do what depends on willing that are inapplicable to abilities to will; some of them offer no analysis of the latter. Moore (1912) is an example with respect to both points. Fara (2008) and Vihvelin (2004) are more recent examples with respect to one or more of these points. Fara acknowledges that the view he offers applies only to abilities to do things that we do as a result of trying to do them; he offers no account of abilities to try to do such things. Vihvelin maintains that an ability to make choices on the basis of reasons is a bundle of dispositions. But whereas she offers an analysis of the dispositions said to constitute abilities to do things that one might choose to do, in her 2004 contribution she offers no analysis of the dispositions said to constitute an ability to choose.

⁸ Chisholm says that "the metaphysical problem of freedom does not concern the *actus imperatus*; it does not concern the question whether we are free to accomplish whatever it

Locke himself rejected the notion that the will might be free. Will, he said, is a power, and freedom is a power; and it's nonsense to think of a power having a power.

Let us grant these premises. Still, just as we can ask whether an agent is free with respect to doing a certain thing that she might will to do, so we can ask whether she is free with respect to *willing* to do that thing. If the latter freedom consists in some power, that power would be (as Locke insisted freedom is) a power *of the agent*, not a power of her will.⁹ Having such a power is at least part of what constitutes having free will.

As long as willing is the kind of thing that can be done freely, it makes sense to ask whether we might be free with respect to doing it.¹⁰ What we'll offer, then, is an understanding of willing on which it makes sense to think that one might freely will one thing or another.

If we can do anything freely, we can *act* freely, or freely *perform actions*. We'll take it that willing is performing an action of some kind. And although one might do something freely without doing it intentionally,¹¹ commonly when one freely *A-s*, one *A-s* intentionally. Paradigmatic instances of what we'll call willing, then, will be instances of intentional action (though not every intentional action is an instance of what we'll call willing).

It bears emphasizing that the understanding of willing we'll propose *isn't* meant to capture what might be expressed when one speaks of someone having the will to do a certain thing, or being willing to do it, or of strength or weakness of will. Our target is something an agent's powers with respect to which might be said to constitute (at least to an important extent)

is that we will or set out to do; it concerns the *actus elicited*, the question whether we are free to will or to set out to do those things that we do will or set out to do" (1966: 23). We allow that it might concern the former as well, though, with Chisholm, we hold that the latter is included.

⁹ Compare Hobart's remark that "in accordance with the genius of language, free will means freedom of persons in willing, just as 'free trade' means freedom of persons (in a certain respect) in trading" (1934: 8).

¹⁰ Locke himself declared at one point that "in respect of willing, a man is not free" (*Essay*, Bk. 2, Ch. 21.22). Although he might be correct, we do not find his argument persuasive. A key premise is that once an action in one's power is proposed to thought, it is "unavoidably necessary" to prefer the doing or the forbearance of that action (*Essay*, Bk. 2, Ch. 21.23). But first, the argument doesn't address the case in which one *doesn't* consider doing the thing in question. Second, when one *does* consider this, one can indeed fail to make up one's mind—fail to choose one way or the other—without having decided not to choose. (On this point, see Kane 1996: 156 and Clarke 2014: 100–4.) Third, even if in some case one must either will the act or will to forbear it, one might be free with respect to which of these one wills.

¹¹ When one attempts something with little chance of success, if one succeeds one's action might be too lucky to count as intentional. It can still count as something done freely. Mele (2006: 25) provides an example.

freedom of the will. We'll use 'will', 'wills', and 'willing' as terms of art for this purpose.

In light of these considerations, we propose to include as instances of willing—or, as we'll say, acts of will—two kinds of thing: decisions, and beginning portions of attempts. We'll first note some things that *don't* count and then describe further these two kinds of thing that do.

Given our purposes here, neither *wanting* nor *intending* to do a certain thing counts as an instance of willing. Wanting and intending are states, not actions. Further, one can want, and even intend, to do a certain thing without so much as beginning to act as one wants or intends.

Nor will *coming to have an intention* always count as an instance of willing. True enough, one can come to have an intention to *A* by deciding (or choosing) to *A*. And a decision to *A* is, we take it, an intentional action. (We'll give our reasons for saying so shortly.) But one can come to have an intention in a way other than by deciding. Deciding settles practical uncertainty, uncertainty about what to do. However, on many occasions when we act, it's perfectly clear what to do; there's no uncertainty about it. When this is so, one can come to intend to do a certain thing, and do it intentionally, without deciding to do it.¹²

Still, we count *deciding* as a kind of willing. Making a decision commonly has a phenomenal quality characteristic of action; it often seems to oneself, at the moment when it's done, that it's something that one is actively doing. We take the appearance not generally to be illusory. Making a decision is actively forming an intention to do a certain thing, and intentionally forming that intention. (We don't say that to intentionally decide to *A*, one must intend to decide to *A*; perhaps an intention to make up one's mind whether to *A*, or something of this sort, will do.) A decision to *A* is an action distinct from *A*-ing, one that can be performed well before, or without ever, doing (or even attempting to do) what one therein decides to do.¹³

¹² An example (from Mele 1992: 231): when Al arrives at his office in the morning, he reaches into his pocket, takes out his keys, and unlocks the door. He does all of this intentionally, and when he does it he intends to do it, and to do it then. The intention to unlock the door right then is something he comes to have on his arrival. But he doesn't make a decision to unlock the door then. Given his standing plans and his awareness that he's arrived at the door, it's perfectly clear to him what to do then. There's no practical question that needs to be settled, and he comes to have the intention in much the same way that he might come by inference to have a belief. *Executing* that intention, *carrying it out*, is an action he performs; coming to have the intention is not. For another example of acquiring an intention without making a decision, see Audi (1993: 64).

¹³ For arguments in support of the view that decisions to act are themselves intentional actions, see McCann (1998b); Mele (2000); and Pink (1996). O'Shaughnessy (2008: 543–47) denies that deciding is an action; for a reply, see Clarke (2010a).

If Al *decides* today to drive to his office tomorrow, then we'll say that he *wills* today to make the drive tomorrow. He actively settles on doing that; he *sets his will* on performing that future action. Similarly, when one decides to do a certain thing right away, the making of the decision is itself something that we'll call an instance of willing, or an act of will.

Another thing that we'll count is a beginning portion of an attempt to do something. By attempts we mean instances of trying. And we employ here a notion of trying on which it can be effortless; some attempts are easy. When you intentionally raise your arm, even in normal circumstances, you've made a successful attempt to raise your arm. And a successful attempt is an attempt, or an instance of trying.¹⁴

Like decisions, attempts are intentional; if you try to *A*, you intentionally try to *A*. Although intentionally trying to *A* might require having some intention, it doesn't require having an intention to try to *A*. When you try to raise your arm, you might simply intend to raise your arm.

We don't say that trying is itself willing. To see why, consider an action such as sinking a putt during a game of golf. It's fruitful to think of such a thing as a process, one that begins when certain changes occur in the agent's head—when her having an intention to putt the ball begins to activate certain motor neurons¹⁵—that continues with the delivery of motor signals to certain muscles, and includes motions of her shoulders, arms, and hands. Philosophers disagree about whether such an action includes such things as the motion of the club, the striking of the ball, the rolling of the ball across the green, and the descent of the ball into the cup.

Suppose that it does. How much of this process consists of the agent's willing to sink the putt? Surely not the ball's falling into the cup; indeed, we might well deny that this occurrence is any part of the attempt.¹⁶ But what about the back and forth motion of the agent's hands? This is part of the attempt, but we decline to count it as part of the agent's act of will.¹⁷ One

¹⁴ Saying that one tried to raise one's arm can implicate that it was a difficult matter, or that one didn't fully expect success. But we distinguish pragmatic infelicity from falsehood; the statement might be nonetheless true.

¹⁵ Brand (1984: 20) suggests this view of when actions begin.

¹⁶ Well before the ball falls into the cup, the agent has ceased to make any movements of her body that might influence whether she sinks the putt. She might at that moment be trading quips with her partner or walking toward her golf bag. She isn't, while doing any such thing, trying to sink the putt. (We assume that if some portion of an agent's *A*-ing is occurring at a given time *t*, then the agent is *A*-ing at *t*.)

¹⁷ O'Shaughnessy (2008) *does* count such motions as included in an agent's willing to do a certain thing.

consideration against doing so is that, while the motion of the hands might be swift or slow, and might describe an arc, it seems wrong to say that a portion of the willing is swift or slow, or describes an arc.

But consider an early part of this process, a beginning portion of the execution of the agent's intention to make the putt. As we suggested, this takes place when the agent's having an intention to putt the ball activates certain motor neurons in her head. This is something that, for our purposes here, we take to be willing. It is an intentional action, an agent's freedom with respect to which might well be thought of as (an aspect of) her freedom of the will.

There will be a string of beginning portions of such a process occupying several moments, as long as the agent's having that intention continues to excite the appropriate motor neurons. Indeed, one's willing to do a certain thing, understood as consisting of beginning portions of the execution of one's intention to do that thing, might go on just about as long as does one's doing that thing, when one's action extends no further than motions of one's body. When you intentionally raise your arm, your willing to raise it occupies about the same time period as does your raising it.

It's sometimes said (e.g., by Mele (2000)) that deciding is a momentary act. If this is correct, should we take the acts of will that we've just identified to be momentary as well? If we do, we should observe that many an action, such as the sinking of the putt or the raising of the arm, involves one or more continuous strings of such momentary willings, each a beginning portion of a continuing attempt. We might instead take the act of will in such a case to be the series of beginning portions of the attempt. We'll couch the subsequent discussion in terms of the former view, though we don't see that anything of importance to our purposes here hinges on which alternative is adopted.¹⁸

We've identified two things that we count as willing: making a decision, and engaging in a beginning portion of an attempt. (There might be other things that should count;¹⁹ we'll confine our attention to these two.) Our

¹⁸ We've said that beginning portions of attempts count as willings. Why not more simply take beginning portions of *actions* to be willings (perhaps except when the actions in question are themselves momentary)? In trying to *A*, one wills to *A*, whereas in *A*-ing one need not will to *A*. Stating the view in terms of trying, we hope, makes it clear *what* is said to be willed on a given occasion.

¹⁹ O'Shaughnessy (2009) argues that we have trying only where it is possible to try but fail, and that in mental actions of certain types (e.g., imagining moving one's arm), there is no such possibility. We then have what he calls "a movement of the will" (we might say, a willing) without an attempt. If he is correct, then we have here a third kind of willing, something additional to the two kinds we've identified in the text.

aim, as we've said, is to identify a kind of action one's freedom with respect to which might be one's freedom with respect to willing. A powers-conception of free will, we maintain, needs to provide an account of the latter, for freedom of the will is, in the first instance, freedom with respect to willing to do this or that.

It's quite common for theorists of free will to focus on decisions (or choices—we take these to be the same). The second sort of thing we've identified shares some features with what some have called volition, which, again, is often a focus of free-will theorists. We think the focus well intentioned, but the sort of willing we've identified differs in important respects from volition.

Theories that appeal to volition commonly take it to be a distinctly mental action with representational content, in the same way in which a decision to *A* has content, representing one's *A*-ing. But engaging in a beginning portion of an attempt to *A* is no more a distinctly mental action than is trying to *A*, and trying to *A*, when one succeeds, is no more a distinctly mental action than is one's *A*-ing. Unlike deciding, the second kind of willing we've described isn't the coming to have or the formation of some psychological attitude. It is, rather, a beginning portion of the execution of some such attitude—a portion of the execution of an intention.

Volitionists often locate volition, in cases of overt, bodily action, causally or temporally (or both) between desiring or intending to do a certain thing and beginning to activate one's motor system.²⁰ What we've called trying isn't such an intermediary, and neither is a beginning portion of an attempt. In cases of overt, bodily action, a beginning of the activation of the motor system *is* a beginning portion of an attempt to do a certain thing; and when one succeeds, it might be a beginning portion of one's doing that thing. We see no need, either in action theory or in an account of free will, for the sort of intermediary between intention and beginning of execution to which volitionists appeal.

We contend that our view captures something that is right about certain volitional accounts: one's freedom with respect to engaging in beginning portions of attempts is a crucial aspect of free will. But we reject the appeal to volition as something in addition to intending, deciding, and trying. Our view thus avoids what we see as questionable aspects of many volitional accounts.

²⁰ Ginet (1990: 9–14) takes volitions to be mental actions that, in cases of bodily action, cause bodily exertion. On his view, a volition need not have any cause or any internal causal structure.

1.2 UP TO YOU

It's a common thought that if you have free will, then at least sometimes when you act, it's up to you whether you do the thing you do on that occasion. Applying this idea to willing, we can say that if on some occasion you're free with respect to willing to *A*, then on that occasion it's up to you whether you then will to *A*.

Where the willing at issue is the making of a certain decision, say, the decision to *B*, if on some occasion you're free with respect to making that decision, then it's up to you whether you then decide to *B*. Where a beginning portion of an attempt to *C* is in question, if on some occasion you're free with respect to that beginning portion of the attempt, then it's up to you whether you then engage in that act of will.

The expression 'it's up to you' is often used in a way that *isn't* concerned with free will. A father might have said to his daughter on some occasion: "You can straighten up your room, or you can stay home; it's up to you." A certain outcome—whether she goes out or stays home—is made to depend on whether she straightens up her room. Whether she does the latter or not will be the difference-maker. The choice is hers. But all this might be so whether anyone has free will or not.

Similarly, if someone has no preference regarding what you do on some occasion, or doesn't wish to express one, or refuses to offer any suggestions or advice or exert any pressure, she might say the matter is up to you. Make the decision without her direction. That's something you can do whether we have free will or not.

But if we have free will, then not only do certain outcomes depend on what we will to do, but also it's sometimes up to us whether we will to do certain things. And its being up to us whether we will to do certain things isn't a matter of no one's offering us advice or direction. Indeed, it can be up to you whether you will to do a certain thing even if someone *is* offering you advice, expressing her preference about what you do, or exerting pressure on you to do a certain thing.

We take this requirement to impose a constraint on any theory of free will that aims to capture our ordinary idea of this phenomenon. And we'll take it as a constraint on construals of powers possession of which by an agent amounts (perhaps with further conditions) to that agent's having free will. It must be the case that having these powers (perhaps together with satisfaction of the further conditions) can suffice for its sometimes being up to an agent whether certain of the powers are exercised.

It might *not* be a requirement of *moral responsibility* that we have such powers. It might be that one can be responsible for doing something even though it wasn't up to oneself on that occasion whether one willed to do

that thing. We take the possibility here to be the possibility that agents might not have to have free will in order to be morally responsible for some of the things they do.²¹

That human agents commonly have such powers is nevertheless, we submit, part of our ordinary conception of our agency. When one is deliberating about whether to *A*, one ordinarily takes it as given that it's up to oneself whether one decides to *A*. One presumes not just that one's decision will be a difference-maker with regard to what happens subsequent to it, but that it's up to oneself whether one makes one decision or another. And one might presume this even if one knows that someone else is pressing one to decide a certain way. It is free will of the sort that is presumed in this kind of case that is our target here.

Expressing what's required for free will in this way—in terms of something's being up to oneself—is preferable to some common alternatives. Sometimes it's said that one has free will only if, at least sometimes when one acts, one can do other than what one actually does. But the statement doesn't clearly indicate what's required. Even when she's sleeping, it's true that Elena can speak French; but she isn't just then free with respect to whether she does so. Similarly, stating what's required in terms of an ability to act is problematic; despite lacking a guitar, one might have a general ability to play that instrument.²² Even putting things in terms of what's within an agent's power might fall short, if, as seems possible, one might have a power and yet it not be up to oneself whether that power is exercised.²³ And freedom with respect to *A*-ing isn't well expressed in terms of having a choice about whether one *A*-s, when one's *A*-ing might itself be one's making a choice (say, to *B*).²⁴

The penultimate point suggests a difficulty for a powers-conception of free will: what must an agent's powers be such that on some occasion it's up to her whether and how they're exercised? We think that this is indeed a difficulty. It's one to which we'll repeatedly return.

²¹ One might put the intended point of Frankfurt (1969) this way (as, indeed, Frankfurt himself does).

²² Modalities expressed by sentences using 'can' or 'able' differ depending on context, and it might be argued that in the context of discussion of free will, stating what's required using these terms is satisfactory. This might be so, but the frequency with which people talk past each other while discussing free will in these terms gives us little confidence that such sentences are generally understood the same by all participants in the discussion.

²³ Van Inwagen (1983: 8–9) employs all three of these terms in explaining what he means in saying that someone has free will. He also often (e.g., 1983: 16, 2008a) uses our favored expression in characterizing free will.

²⁴ A much discussed version of the Consequence Argument (van Inwagen 1983: 93–104) is couched in terms of having a choice about whether something or other is so.

1.3 POWERS

We've offered a view of what willing is, one that can be employed in an account of freedom with respect to willing; and we've identified a crucial feature of the kind of freedom with respect to willing that is at issue with free will. What about powers?

Powers are a class of properties including dispositions, capacities, abilities, tendencies, susceptibilities, and liabilities. We include here both what Locke called active powers and what he called passive powers. Perhaps the common feature of such properties is what's sometimes called *directedness*: each is directed at some manifestation (or perhaps some plurality of manifestations).²⁵ A manifestation of fragility is breaking, one of narcolepsy is falling asleep at apparently random moments.

Powers differ in several respects, and proponents of powers-conceptions of free will thus face some decisions about the kind of powers they'll invoke. Some powers are what we'll call *causal dispositions*. Each of these has one or more characteristic stimuli in response to which the disposition is manifested, when it is; and the stimuli are among the causes of the manifestation. Fragility is such a power. Striking is a characteristic stimulus of it, and when a struck object manifests its fragility, the striking of that object is a cause of the object's breaking.

Several recent powers-conceptions have invoked causal dispositions. We'll call views of this sort *dispositional accounts of free will*. While we think the recent proposals have fallen short, we want to see how far one can go with such a view. Most of what we say in the remainder of the paper is an exploration of this question.

But different kinds of powers might be invoked. And we think it important to acknowledge explicitly the choice one makes on this matter.

On some views of free will, when an agent freely makes a certain decision, the *agent* causes something, such as her coming to have a certain intention, and the agent's causing that thing isn't causation by any occurrence or state. It's causation by an enduring substance, which, on this kind of account, is what a rational agent is.²⁶ If we take it that an agent who causes something has a *power* to cause, then on this view having free will requires having a causal power, albeit one whose manifestation is something caused by the agent. Such a power might have characteristic stimuli that are somehow relevant to its manifestation, but, it's usually said, any

²⁵ The term 'directedness' is used by Molnar (2003: 60) to characterize what he calls powers or dispositions.

²⁶ For recent views of this sort, see Clarke (1993); Jacobs and O'Connor (2013); Mawson (2011: ch. 5); O'Connor (2000); and Pereboom (2001: ch. 2). Earlier versions are advanced in Chisholm (1966) and Taylor (1966).

such stimulus isn't a cause of the manifestation (or of the agent's causing that manifestation).

Other powers-conceptions appeal to spontaneous powers. Free will aside, there seem to be powers whose manifestations are independent of any stimulus. The instability of certain particles and elements is manifested in spontaneous decay. There seems to be no stimulus needed, and none relevant, to whether the kind of decay in question occurs. And some theorists maintain that having free will is having a spontaneous power, though one that, unlike the instability just described, is a rational power, one that is exercised for reasons.²⁷

We're doubtful that either of these alternative approaches improves on an account appealing to causal dispositions. Toward the end of the paper we'll briefly explain the ground of our doubt. But our main task from here forward is to offer some proposals concerning how a dispositional account might best go.

One difficulty for any such effort stems from the requirement, identified earlier, that if one has free will, then on some occasions it's up to oneself whether one wills this or that. Setting aside for the moment the possibility of indeterminism, whether a causal disposition is manifested on some occasion is just a matter of whether sufficient causal antecedents occur. When a fragile glass is struck, it breaks—or it doesn't—depending on how it's struck, how fragile it is, and perhaps other features of the glass and the situation. Plainly it isn't up to the glass whether it breaks. And now, if an agent's powers are all of this general sort—if they're all causal dispositions—then whether they're manifested on some occasion is likewise just a matter of whether sufficient causal antecedents occur. It's then hard to see how it can be up to an agent on some occasion whether some power to will that she possesses is manifested.²⁸ Among the things we'll want to see is what a proponent of a dispositional account of free will can say in response to this difficulty. (Later we'll drop the assumption of determinism and consider whether doing this helps.)

²⁷ Lowe (2008: chs. 7–8) advances an account of free will that appeals to spontaneous powers. Goetz (2008: ch. 2) and Pink (2011) similarly appeal to powers whose manifestations are said to be uncaused. McCann (1998a: 180) emphasizes the spontaneity of exercises of free will, which he takes to be uncaused, though he doesn't couch his account in terms of powers.

²⁸ On the basis of considerations similar to those presented here, Taylor denies that the "can" of human agency expresses "the idea of a causal relationship between different events or states" (1960: 81–82); and van Inwagen contrasts an agent's "power to act" with a "causal power or capacity" (1983: 11).

1.4 A POWER TO TRY

Imagine a young child who sees a shiny object across the room, crawls over to the object, and grasps it.²⁹ The child, we may suppose, lacks certain rational capacities that would be needed for her to have free will. But we shouldn't deny altogether that she's an agent. She has, and on this occasion exercises, many of the powers that are distinctive of agency. Her behavior is motivated, goal-directed, and guided by her in light of her perception of her environment.

The child exercises a power to crawl over and get the shiny object. In crawling to get the object, she carries out an attempt to do that—she tries to crawl and get it—and she exercises a power to make that attempt. (Recall that trying to *A* doesn't require intending to try to *A*; when trying to *A*, one might simply intend to *A*.)

When the child makes the attempt, there occur a series of beginning portions of that attempt. Employing the notion of willing developed earlier, we may say that the child wills to crawl and get the object. (Again, she might so will without intending to will anything, and indeed without having a concept of willing.)³⁰ She exercises—and hence has—a power to so will.

The child can have this power even if, as we may presume, it isn't up to her whether she exercises it on this or any other occasion. One can have a power to will without having free will, without having a power to freely will.

We might manage an account of a power to will to do a certain thing if we set our sights lower than free will, on something that the young child has despite lacking free will. When one is trying to *A*, one is executing an intention. To be engaged in such a thing is to be willing to do something; it is to manifest a power to will.

Consider a construal of a power to engage in a beginning portion of an attempt to *A* as a power to engage in such a thing in response to coming to have an intention with relevant content—for example, an intention to *A*. This proposal allows us to see a power to will as a causal disposition, one whose manifestation has among its causes a characteristic stimulus.³¹ One

²⁹ We borrow the example from Vihvelin (2013: 178). The present paper began as, in part, a critique of that work but has evolved into something rather different—an effort to see if we can ourselves produce a view that avoids the problems we see in those of others.

³⁰ If the child lacks a concept of trying, or a concept of willing, might it be that her trying or willing to go and get the shiny object isn't intentional? (Stephen Kearns raised this question.) We allow that this might be so. Then our claim that acts of will are intentional actions would need to be qualified, perhaps as applying only to agents who can freely will.

³¹ Things are a bit more complicated if we see actions (as some action theorists do), including willings, as causings. The child's willing to crawl might then be a series of causings—causings of beginning portions of her attempt by her intending to crawl. Let

advantage of the proposal is that we have a reasonably good understanding of causal dispositions. And it might be thought that, in any case, an account of agency must see intentions (or acquisitions of intentions) as causes, for (one might think—and we're inclined to think) that's the way to understand the implementation (the carrying out) of an intention.

In any event, having powers to engage in beginning portions of attempts to do various things apparently requires having a host of other powers. For one thing, one must have a power to come to have intentions (or functionally similar executive states). Intentions are themselves commonly motivated in light of beliefs. Hence, it seems, in order to have powers to will, one must have powers to come to have cognitive and motivational states.

A power to engage in a beginning portion of an attempt to do some *specific* thing would require, it seems, having some specific motivational and conceptual capacities. An agent who lacks a power to become motivated to *A*, or motivated to try to *A*, or to have any other relevant motivation, would lack a power to engage in a beginning portion of an attempt to *A*. Likewise for an agent who lacks a power to come to think of doing a certain thing.³²

Even setting aside the problem of free will, then, a power to will includes or depends on a variety of powers, many of them not powers to do things intentionally. We have no proposal concerning these further powers beyond the suggestion that they, like a power to will, might be seen as causal dispositions, with characteristic stimuli that are among the causes of their manifestations, when the manifestations occur.

1.5 RATIONAL POWERS

Our powers to will are, even if those of a young child aren't, *rational* powers. And if we're eventually to have a conception of free will in terms of agential powers, it will have to include a conception of rational powers. No agent with free will entirely lacks the latter.

Among the relevant rational powers are powers to come rationally to believe, desire, and intend. If these are powers to do various things in response to various stimuli, the relevant stimuli might include a kind of

us say that what is caused in a causing is a causing-result. Then intending will cause, not willings, but willing-results. A power to will is still seen as a causal power, even if the respect in which it is so must be understood a bit differently.

³² A medieval knight would lack a power to try to find an iPhone, in part because he lacks a power to come to have an intention to search for an iPhone. And he lacks the latter power because he lacks a power to think of doing such a thing. He could very well come to have that power—someone could tell him about iPhones. But that's to say that he has a potential (a power) to acquire a power that he in fact lacks.