MARK SCHROEDER

EXPLAINING THE REASONS WE SHARE

Explanation and Expression in Ethics, Volume 1

OXFORD

Explaining the Reasons We Share

Explaining the Reasons We Share

Explanation and Expression in Ethics, Volume 1

Mark Schroeder

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

First Edition published in 2014

Impression: 1

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: 2013954861

ISBN 978–0–19–871380–7

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

Dedicated with Gratitude to Todd and Dawn Nelson

Preface

That part of philosophy which is concerned with understanding the nature of moral facts, thought, language, and knowledge, or more generally of *normative* facts, thought, language, and knowledge, of which morality is just a special case, is known as metaethics. On a typical way of thinking about metaethics, it is a branch of applied philosophy. On this view, the study of moral facts is applied metaphysics, the study of moral thought is applied philosophy of mind, the study of moral language is applied philosophy of language, and the study of moral knowledge is applied epistemology. The unifying theme of this volume of essays and its sister volume, *Expressing Our Attitudes*, in contrast, is that metaethics is much too intimately tied up both with normative inquiry and with fundamental questions in each of these 'core' areas of philosophy for this picture to be correct. In *Expressing Our Attitudes* I argue for a particular picture of how deeply one prominent strategy for understanding moral thought and language pushes us into the project of completely re-thinking the nature of thought and language more generally. And in this volume I argue that in its most ambitiously general and explanatory forms, normative inquiry itself is thoroughly enmeshed in metaethical inquiry.

Normative inquiry in its most ambitiously general form aspires to tell us not only what is good, bad, right, wrong, just, and otherwise, but *why*. In other words, it purports to be explanatory. The essays in this book are all about the kind of explanation that normative inquiry aims to provide, and about the metaethical commitments that we incur in offering such explanations. They both document the development of my views about reasons, reduction, and normative explanation over a period of about twelve years, and collectively argue for a particular conclusion: that normative inquiry can offer perfectly general explanations only if it is *reductive*, and that reductive normative explanation is a powerful and fruitful project. The essays selected for the volume have been chosen more with an eye to advancing my cumulative case for this conclusion, than for the representativeness of the body of my work on similar topics. I have left out closely related papers, for example, on moral particularism, reasons and rationality, and on the relationship between value, reasons, and obligation.

Readers who are already familiar with some of my other work may recognize one important theme in common between the emphasis of these essays and my preoccupations in *Slaves of the Passions*—a concern for the special need to explain not just why we have the reasons or obligations that we do, but specifically why we *share* the reasons or obligations that we seem to share. Without replacing what I have written there, it is my hope that these essays put the pressure to take on this explanatory burden in a broader context. They also advocate a broader conception of what a successful such explanation might look like, and I hope that they give us at least a helpful way of thinking about what kind of diversity of approaches might be needed or possible.

The essays included in the volume cover a roughly twelve-year span, and in some cases make different choices of terminology, as well as adopting differences in both emphasis and approach. As I explain in the introduction, I no longer believe quite every idea I espouse in each of the essays, but I do endorse what I think are their most important arguments and themes. In editing them for inclusion in this volume, I have elected to preserve their content as previously published as closely as reasonable, editing only for remaining typographical errors, consistency of layout and to update bibliographical entries, where appropriate.

Roughly half of the work represented in this volume was conceived and executed during the last seven years, while USC's Mudd Hall of Philosophy, represented in the jacket art, has been my professional home. My colleagues in the USC School of Philosophy believed in me before others did, and colleagues, administrators, and students at USC have made it an exceptionally rewarding place to do philosophy.

Far too many people have assisted me in the work assembled in this volume for me to accurately be able to thank them all. But special thanks are due, among others, to Karen Bennett, Michael Bratman, John Broome, Krister Bykvist, David Copp, Stephen Darwall, Jamie Dreier, Billy Dunaway, Ant Eagle, David Enoch, Steve Finlay, Maurice Goldsmith, Pete Graham, Gilbert Harman, Liz Harman, John Hawthorne, Tom Hill, Zena Hitz, Brad Hooker, Nadeem Hussain, Aaron James, Scott James, Robert Johnson, Mark Johnston, Jeff King, Sari Kisilevski, Shieva Kleinschmidt, Joshua Knobe, Niko Kolodny, Barry Lam, Doug Lavin, Stephan Leuenberger, Alida Liberman, Errol Lord, Mike McGlone, Tristram McPherson, Michael Morreau, Mark Murphy, Shyam Nair, Alasdair Norcross, Philip Pettit, Jim Pryor, Joseph Raz, Baron Reed, Gideon Rosen, Jake Ross, Gillian Russell, Geoff Sayre-McCord, Johannes Schmitt, Andrew Sepielli, Kieran Setiya, Russ Shafer-Landau, Rob Shaver, Brett Sherman, Sam Shpall, Walter Sinnott-Armstrong, Michael Smith, Justin Snedegar, Scott Soames, David Sobel, Nic Southwood, Jeff Speaks, Bart Streumer, David Sussman, Valerie Tiberius, Jens Timmerman, Mark Timmons, Jim Van Cleve, Mark van Roojen, Jonathan Way, and Ralph Wedgwood, as well as to a number of blind referees whose names I do not know.

Portions of the work appearing in this volume have benefited from audiences at Bowling Green State University, Georgetown University, New York University, Northwestern University, Princeton University (on several occasions), Rice University, the University of Iowa, the University of Maryland, the University of Reading, the Berkeley/Stanford Graduate Conference in Philosophy, the Columbia/NYU Graduate Conference in Philosophy, the Carolina Metaethics Workshop, the College Park Conference on Practical Rationality, the Oxford University Moral Philosophy Colloquium, and the second annual Wisconsin Metaethics Workshop.

All but one of the papers included here have been previously published, and I'm grateful to each of their respective publishers for the opportunity to reprint them here. I'm especially grateful to Johannes Schmitt, my co-author of 'Supervenience Arguments Under Relaxed Assumptions' for his permission to publish our paper under a collection of 'my' work. The volume would be far less unified without Peter

Momtchiloff's invaluable guidance in selecting papers for inclusion, as well as the help of his readers for Oxford University Press, and Caleb Perl's assistance was instrumental in preparing and finalizing the manuscript. Sacrifices made by Dawn, Todd, and Madeleine Nelson were what made it possible for me to bring this work to completion, and I owe them an incalculable debt of gratitude. But my biggest debts are to Maria Nelson, who makes me a better me.

Acknowledgments

All but one of the papers published in this volume have been published previously or are committed to publication. I'm thankful to the following for the opportunity to reprint them here, together.

'Cudworth and Normative Explanations'. *Journal of Ethics and Social Philosophy* 1(3), www.jesp.org (2005).

'Reasons and Agent-Neutrality'. *Philosophical Studies* 135(2): 279–306 (August 2007).

'The Humean Theory of Reasons'. *Oxford Studies in Metaethics* 2: 195–219 (April 2007).

'What Matters about Metaethics?' Forthcoming in *Parfit's Defense of Objectivity in Ethics and Practical Reasoning*, forthcoming from OUP, edited by Peter Singer.

'Supervenience Arguments under Relaxed Assumptions'. With Johannes Schmitt. *Philosophical Studies* 155(1): 133–160 (August 2011).

'The Price of Supervenience'. Not previously published.

'The Scope of Instrumental Reason'. *Philosophical Perspectives* 18 (Ethics): 337–364 (November 2004).

'Means–End Coherence, Stringency, and Subjective Reasons'. *Philosophical Studies* 143(2): 223–248 (March 2009).

'The Hypothetical Imperative?' *Australasian Journal of Philosophy* 83(3): 357–372 (September 2005).

'Hypothetical Imperatives, Scope, and Jurisdiction'. Forthcoming in a volume of new essays honoring Thomas E. Hill, Jr., forthcoming from OUP, edited by Robert Johnson and Mark Timmons.

'Scope for Rational Autonomy'. *Philosophical Issues* 23 (Epistemic Agency): 297–310 (August 2013).

Contents

Introduction 1

Part 1

1. Cudworth and Normative Explanations 19
2. Reasons and Agent-Neutrality 42
3. The Humean Theory of Reasons 60

Part 2

4. What Matters About Metaethics? 83
5. Supervenience Arguments Under Relaxed Assumptions 96
6. The Price of Supervenience 124

Part 3

7. The Scope of Instrumental Reason 147
8. Means–End Coherence, Stringency, and Subjective Reasons 173

Part 4

9. The Hypothetical Imperative? 201
10. Hypothetical Imperatives, Scope, and Jurisdiction 216
11. Scope for Rational Autonomy 227

References 241
Index 247

Introduction

Sasha takes a job working as a server at Bob's Diner. Her job is to take care of her customers' wishes, and she wants to do well at her job. So we may assume that she ought to take care of her customers' wishes. Tonight she is responsible for tables 15 through 21, so the customers sitting at tables 15 through 21 are hers. So she ought to take care of the wishes of the customers at tables 15 through 21—and we know why: it is because they are her customers, and she ought to take care of her customers' wishes. Todd is a customer sitting at table 17. So she ought to take care of Todd's wishes. And again, we know why: it is because Todd is sitting at table 17, and she ought to take care of the wishes of the customers sitting at tables 15 through 21. Todd wishes to order an untossed cobb salad with no avocado, with dressing on the side. So Sasha ought to bring him an untossed cobb salad with no avocado, with dressing on the side. As before, we know why: it is because this is what he wishes, and she ought to take care of his wishes.

Examples like this are simple and natural, and can easily be multiplied. They are cases in which specific obligations are explained in terms of more general ones. It's because of Sasha's general obligation to take care of her customers that tonight she has an obligation to take care of the customers at tables 15 through 21, which is in turn why she has an obligation to take care of Todd, which is in turn why she has an obligation to bring him an untossed cobb salad with no avocado, with dressing on the side. In each case the more specific obligation doesn't just *follow* from the more general one, but it is intuitively *explained* by it. Sasha's specific obligations are *inherited* from her more general obligations.

In the same way that one agent's specific obligations are often explained by her more general obligations, individual agents' obligations are often explained by shared obligations. For example, Logan, who is also a server at Bob's Diner, also ought to take care of his customers. But tonight Logan is responsible for tables 8 through 14. So though both Sasha and Logan ought to take care of their customers' wishes, only Sasha ought to take care of the wishes of the customers at tables 15 through 21, and for this reason only Sasha ought to take care of Todd's wishes, and hence to bring him the untossed cobb salad, hold the avocado. Whenever an individual agent's specific obligations are explained by general obligations, together with specific features of her circumstances,

we may also get an explanation of how her distinctive individual obligations derive from obligations that are shared.

When most philosophers refer to 'moral explanations', what they have in mind are explanations of *non*-moral facts, by appeal to moral facts as *explanans*—the things which do the explaining. In this sense, moral explanations have been made famous by the much-discussed dispute between Gilbert Harman and Nicholas Sturgeon. But when I use 'moral explanation' in this volume, what I have in mind are explanations *of* moral facts as *explananda*, as in the cases involving Sasha's obligations. On a natural conception, the central point of normative moral theorizing is to offer moral explanations in this sense—to say not only *what* people ought to do, but *why*. What the case of Sasha illustrates is that we know a lot about how moral explanations *often* work. In many cases, specific obligations are explained by being subsumed to and thereby inherited from more general ones. Because such cases are typical or standard, I say that such moral explanations follow the *Standard Model* for normative explanations.

The Standard Model for normative explanations applies to more than obligations. It applies to reasons, for example, in exactly the same way. We know that Sasha has a reason to take care of the wishes of the customers at tables 15 through 21, and we know why—it is because she has a reason to take care of her customers' wishes, and given her circumstances, the customers at tables 15 through 21 are hers. The Standard Model also applies to explanations of what is good. When we agree that money is good because it lets you buy things—and note that this is a good explanation precisely because buying things is good—we are again subsuming more specific claims about what is good under more general claims, by way of inheritance—money inherits what goodness it has from the goodness of what we can do with it. The structure of such explanations of what is good is so familiar that we have a standard piece of philosophical terminology for it: when something's goodness is explained in this way, we say that it is only *instrumentally* good, and when we get to something whose goodness cannot be further explained in this way, we say that it is *intrinsically* good.

This book is about the power and limits of Standard Model explanations in moral philosophy and the theory of rationality—of explanations which subsume the specific under the general, and thereby the individual under the shared, by taking the specific obligations of individuals in particular circumstances to be subsumed to, or inherited from, more general obligations which transcend those circumstances. One of the central themes of the volume is that if we restrict ourselves, in doing explanatory normative theorizing, to subsumptive explanations of this kind, then we leave too much unexplained. In order to defend powerful and perfectly general explanatory normative theories, we must accept a different kind of moral explanation—one that appeals to *reductive* claims about what it is for something to be wrong, or an obligation, or a reason. And once we realize this, we need to think differently both about the relationship between metaethics and substantive normative ethical theory, and about what stands most in need of explanation within normative theory.

1 The Standard Model Theory

According to the primary stalking horse of this volume, the *only* moral explanations that we can give follow the Standard Model: they are subsumptive in nature, subsuming specific obligations in context to more general obligations, by appeal to specific features of the agent's circumstances. Outside of my own work, I know of no clear statements of this theory in contemporary ethical theory, but evidence of its effects are everywhere I look. The Standard Model Theory offers a powerful and compelling framework for thinking about a wide range of issues in explanatory moral theory. Many of the essays in this volume are devoted to illustrating how this theory is worthy of serious respect and careful attention, but arguing that it is false.

There are two ways in which we can think of the Standard Model Theory as being motivated. On the first, it is simply a natural generalization from cases like Sasha's. Seeing that cases like Sasha's work in this way, we infer that this is how moral explanations work, in general. According to a more sophisticated motivation, however, we must eliminate alternative conceptions of how moral explanations might work. For example, they can't be causal explanations, because moral facts don't seem to be caused in the same kind of way that ordinary, non-moral events are caused. Once we eliminate each alternative conception of how moral explanations might work, in this way, the Standard Model is what we have left.

Outside of ethics, one important kind of non-causal explanation explains why something is the case by appeal to claims about what it *is* for it to be the case. For example, when we explain why something is a triangle by adverting to the fact that it has three sides, this explanation works because that is what it is for something to be a triangle. It is natural to call such explanations *constitutive*, because they tell us *what it is* for something to be the case; another natural name for them is *reductive*, because such constitutive claims reduce or analyze something in other terms. Could there be constitutive explanations of moral facts? Such explanations would tell us why someone ought to do something by appeal to what it is for it to be the case that someone ought to do something. They would appeal to an analysis of what someone ought to do. If the Standard Model Theory is to be correct, there must be no such explanations in ethics.

However, if the Standard Model Theory is correct, then there are sharp limits on the possibility of perfectly general explanatory moral theorizing. The reason why is simple: if all explanations of why someone ought to do something advert to something more general that she ought to do, then there can be no true claims of the form, 'whenever someone ought to do something, that is because P'. Any such explanation would need to advert to a more general thing that people ought to do, and that claim could not fall under the scope of the explanatory theory, on pain of circularity. So if the Standard Model Theory is correct, then the best that explanatory moral theory can do, is to provide us with a list of the most basic obligations that we share. I call this argument the *Cudworthy Argument*, both because it is due in essentials to Ralph Cudworth's *Treatise Concerning Eternal and Immutable Morality*, and because I think

it is well-worth chewing on again, and the central paper in this volume, 'Cudworth and Normative Explanations', is devoted to exploring this argument and its consequences.

The Cudworthy argument illustrates the power of the Standard Model Theory in a forceful and illuminating way. It sheds light, for example, on why both the first and second waves of British intuitionism, including both figures like Clark, Balguy, and Price, and figures like Moore, Prichard, Broad, and Ross, were characterized by the clustering of such apparently unrelated theses as non-reductivism, pluralism, and rational intuitionism. Price's appeals to the Cudworthy argument in his *Review of the Principal Questions in Morals* make these connections the most explicit: because all moral explanations are subsumptive, everyone must appeal to a basic list of intrinsic moral obligations, so since the only difference between a list of one and a list of three or seven is in degree, rather than kind, we should expect pluralism to result. Since differences between any two agents' obligations, and between obligations at different times or in different possible circumstances can always be explained, our list of basic obligations will include only obligations that are necessary and shared by everyone—so we will have to learn about them in whatever way we learn about other necessary truths. And since reductive explanations would be an alternative to the Standard Model, we know that there can't be any reductions, either.

2 The Challenge of Shared Reasons

If the Standard Model Theory is correct, then differences between agents' reasons or obligations can be explained only by adverting to a reason or obligation that they share, together with a differentiating feature of their circumstances which explains why one, but not the other, must do this thing in order to fulfill their shared reason or obligation. So if the Standard Model Theory is correct, and all differences in reasons can be explained, and there is no infinite regress of explanations, then all reasons must ultimately derive from reasons that are shared by necessarily everyone.

This is a truly powerful idea. If it is correct, then it is literally incoherent to be skeptical about morality on the grounds that it requires reasons that are shared by everyone, no matter what they are like. But one of the most prevalent causes for doubt about the objectivity of morality over the last few decades has been doubt precisely over this question: whether there are any reasons that apply to anyone, no matter what they are like. Puzzlement over whether there could be any such reasons is what motivated Gilbert Harman's moral relativism, what led the Philippa Foot of 'Morality as a System of Hypothetical Imperatives' to deny that the objectivity of moral requirements requires that they give us reasons, and is the strand of Mackie's queerness argument for the error theory championed prominently by Richard Joyce. None of these theorists is puzzled how there could be reasons—they are puzzled how there could be reasons that necessarily everyone shares. Yet according to the natural consequence of the Standard Model Theory that we have just been discussing, there *can't* be any reasons at all unless there are reasons shared by necessarily everyone.

This view is so strong that even Kant, champion of the categoricity and universality of morality as he was, was committed to rejecting it. Kant held, of course, that hypothetical imperatives are less philosophically problematic than categorical imperatives, on the grounds that their existence can be established by analytic means, whereas the possibility of a categorical imperative requires a synthetic argument. Exhibiting the intuitive allure of Standard Model reasoning, Jean Hampton takes him to task over precisely this claim:

> Kant's position on the nature of hypothetical imperatives must be construed (contra his explicit wishes) such that understanding the bindingness of a hypothetical imperative is no easier than understanding the bindingness of a categorical imperative. My interpretation cannot save Kant's belief that the former is more straightforward than the latter; indeed, my argument is that Kant's belief is wrong. The **only** way to analyze Kant's analyticity claim is to do so in a way that locates in hypothetical imperatives the same mysterious objective authority that attends the categorical imperative. Even more strikingly, I have argued that the force of hypothetical imperatives is dependent on, and *is at least in part constituted by*, the force of some antecedent categorical imperative that is in part definitive of instrumental rationality. [Hampton 1998, 165–166]

And Hampton is far from the only important writer in recent decades to be compelled by Standard Model reasoning. This same sort of claim lies behind, for example, Michael Smith's striking 'Anti-Humean' claim in *The Moral Problem* that unless there is convergence in the desires of every agent under the conditions of complete information and ideal reflection, there cannot be any reasons for action at all.

A little reflection makes it somewhat puzzling how the Standard Model Theorist could be right on this score. Though it follows from the fact that Barry and Shanna both have reasons to take care of Darcy that Shanna has a reason to take care of Darcy, Shanna doesn't have her reason *because* Barry and Shanna both have their reasons; on the contrary, they both have their reasons, in part, because Shanna has hers. In general, conjuncts are not true in virtue of the truth of true conjunctions in which they figure; on the contrary, they help to make the conjunctions true. The same goes for conjunctions with more than two conjuncts, and generalizing, for the universal quantifier. If everyone on Shanna's block has a reason to take care of Darcy, that is in part because Shanna does. It's not true that Shanna does in virtue of the fact that everyone on her block does. Universal generalizations are made true (at least in part) by the collection of their instances, rather than the other way around. So it's puzzling, on the face of it, why any universal generalization of this kind would need to be true, in order for there to be any true instances at all.

'Reasons and Agent-Neutrality' offers a diagnosis of how it could be that the existence of any reasons depends on the existence of reasons that are necessarily shared. It is that on this view, some considerations are reasons for everyone to do something in virtue of being reasons to do it *simpliciter*. In other words, the three-place relation, *R is a reason for X to do A* is to be explained in terms of a more basic, two-place relation, *R is a reason to do A* which has no place for an agent at all. Borrowing a piece of terminology that in retrospect I probably should have left alone, I call the former reason

relation 'agent-relational' and the latter 'agent-neutral'. It turns out that the defensibility of this view about the priority of the agent-relational and agent-neutral 'reason' relations turns on the defensibility of the Standard Model Theory for reasons, and so the ideas are in fact closely tied together.

In the same paper, I also offer a direct argument against the Standard Model Theory, and defend an alternative analysis of apparent agent-neutral reason talk in terms of the agent-relational 'reason' relation. The direct argument that I give is similar to that in *Slaves of the Passions*. Its basic thought is simple—it is harder than you might think to say exactly what there must be agent-neutral reason to do, in order to explain the full diversity of agent-relational reasons had by some individual or other. And it is even harder, once we have these actions in mind, to say what the consideration is, which is an agent-neutral reason to do them.

'The Humean Theory of Reasons' focuses again on the implications of the fact that some reasons seem to be shared, while others are reasons only for some people. It offers a methodological argument that under these circumstances, we will learn more about what explains reasons by focusing on what the *unshared* reasons have in common, than by focusing on what the *shared* reasons have in common. The upshot of the paper—a theme I explored in a related way in *Slaves of the Passions*—is that there is something structurally in need of explanation about reasons that are shared, and in particular about reasons that are necessarily shared. If there are such reasons, it must be because of a convergence in the circumstances that make such considerations reasons for each and every agent, individually. The universal generalization is true because each of its instances is, not conversely.

If each of these two papers is on the right track, then Kant *is* right that there is something at least in principle more in need of philosophical support and explanation about the existence of universally shared obligations or reasons. Whether we accept this conclusion or not, in turn, should go hand-in-hand with whether we reject the Standard Model Theory for normative reasons. The Standard Model therefore has implications not only for whether general explanatory normative theory implicates metaethics, but for the explanatory burdens of such theory, and for what dangers may lurk in the form of the kinds of philosophical puzzles that motivate skeptical metaethical conclusions.

3 Reduction in Metaethics

It is sufficient to reject the Standard Model, to accept analyses of normative concepts—claims about what it takes for something to be a reason, or good, or what someone ought to do. Some analyses—such as Moore's analysis of 'right' in terms of 'good'—are not what we would characterize as 'reductive' in the sense that gets the most attention in metaethical inquiry. These analyses account for one normative concept or property in terms of another, and so they stay within the normative or evaluative 'family' of concepts. Some philosophical resistance to reduction—including Moore's own, on the

view he was led to take later in his career—is resistance to analysis of normative concepts in general.

But the most contemporarily prevalent sort of philosophical resistance to reduction is more concerned about 'boundary-crossing' reductions—analyses of normative concepts in wholly non-normative terms. Such analyses would 'take us outside' of the realm of the normative, and are ruled out by the intuition that normative concepts are 'just too different' from non-normative concepts, in order to be analyzed in this way. This 'just too different' resistance to boundary-crossing reductions fits better with the pairing of Moore's selective application of open-question reasoning to 'good' with his allowance that 'right' may be easily analyzed in terms of 'good'.

Though he appears to accept something very much like the Standard Model Theory for 'good', because he analyzes 'right' in terms of 'good', Moore is not committed to the Standard Model Theory for 'right'. In contrast to Sidgwick, there is no basic, underived obligation to maximize the good, on Moore's theory; there is only the fact that no action counts as 'right' unless it maximizes the good. So there is no puzzle about why Moore is a monist about obligation but a pluralist about value—in the case of the good, he is committed to some basic list of what is intrinsically good, and so the question of pluralism is only a question of how many things are on the list. But in the case of the right, there is no basic list of what is intrinsically right at all—only an analysis of what it is to be right. And there is no puzzle about why there would be only one analysis of 'right'.

Moore is far from the only theorist to split his allegiances between Standard Model and constitutive styles of moral explanation, depending on which moral or normative concept is under discussion. So long as some normative concepts reduce to others, we can offer constitutive explanations of *those* normative facts. But once we get to the *basic* normative concept that we appeal to in our analysis of all of the others, we can only continue with constitutive explanations if we cross the normative/non-normative boundary and give an analysis of that concept or property in non-normative terms. And this is where many balk, who are otherwise, like the early Moore, perfectly ready to countenance constitutive explanations of at least some normative facts.

The essays in Part 2 take up the grounds for so balking. 'What Matters About Metaethics?' addresses Derek Parfit's extreme pessimism about the implications of the reducibility of the normative to the non-normative. Given my perspective from Part 1, once we appreciate the implications of the Standard Model Theory, we see that explanatory moral theorizing can aspire to perfect generality only if we reject the Standard Model. And constitutive explanations are the most promising alternative model for moral explanations. So given this picture, reductive analyses of the basic normative concepts or properties in non-normative terms will simply be part and parcel of our explicit commitments within the most engaged and ambitious kind of first-order normative theorizing. In contrast, Parfit sees reductive claims in metaethics as only motivated out of some obscure kind of skepticism about normativity itself.

My aim in this essay is simply to lower the stakes on discussions of reductivism in ethics. The only way that we could learn that any reduction of the normative to the

non-normative is true, I argue, is if that reduction vindicates and explains what we really think is important. So there is no way that it could follow from the truth of the reductive theory that everything we think is important turns out to be for naught. The reductive theory may be false, but the question of whether it is so or not should not be so fraught; it should simply be one for substantive normative theorizing of the kind I aimed to do in *Slaves of the Passions*.

But though it is true that I think it follows from my arguments in Part 1 that the most ambitious kind of explanatory first-order normative theorizing commits us to aspiring to find reductions of the basic normative concepts or properties in non-normative terms, I also think that if any such reductive analysis of the normative in terms of the non-normative is true, then that will have explanatory benefits with respect to traditional puzzles of metaethics. In particular, I think it is likely to make it easier to explain how we could have moral knowledge. And most forcefully, in my view, it promises to explain how there could be the kinds of necessary connection between the normative and the non-normative that are entailed by compelling *supervenience* theses.

In general, supervenience theses tell us that there can be no difference of one kind—the *supervenient* kind—without some difference of another kind—the *subvenient* kind. Supervenience is not itself a dependence relation—any domain of necessary truths supervenes trivially on any other domain, without depending on it in any way—but when one domain of properties or truths supervenes non-trivially on another, that is generally evidence of a kind of dependence between them. The single thing which I have found personally must puzzling about non-reductive forms of realism in metaethics, is how they are to explain the necessary connections between the normative and the non-normative that we observe and which are codified by various compelling supervenience theses.

Metaethical non-reductivists typically express either a lack of anxiety about how there could be these necessary connections between what they seem to be claiming are really distinct existences, or appeal to 'partners in crime' of one sort or another. But I know of two serious strategies for addressing this kind of modal challenge, on behalf of the non-reductive realist. One is due to Ralph Wedgwood, and one has been recently advocated by T. M. Scanlon. The next two essays, 'Supervenience Arguments Under Relaxed Assumptions' and 'The Price of Supervenience' take up each of these strategies, respectively.

'Supervenience Arguments Under Relaxed Assumptions', co-authored with Johannes Schmitt, explores what happens to a wide variety of arguments about supervenience when we relax the strict assumption that metaphysical necessity is governed by the 'S5' axiom, which says that whatever is possibly necessary, is necessary—or put differently, that whatever is possible, is necessarily possible. Whereas my own modal qualms about non-reductive moral realism turn on the question of whether it can *explain* the necessary connections required by supervenience, others have argued that supervenience actually *entails* that non-reductive moral realism is false. It turns out that many of the interesting modal arguments about supervenience rely on assuming

that the 'S5' axiom is true. Schmitt and I pull apart which arguments require which assumptions about modal logic, in order to shed light on the plausibility and commitments of Ralph Wedgwood's fascinating and resourceful strategy for responding to the modal challenges.

Wedgwood's core idea, which allows him to offer responses to both direct and explanatory supervenience arguments against non-reductive normative realism, is to reject the S5 axiom. On Wedgwood's picture, not only can something be possible without being necessarily possible, but something can be possibly possible without being possible, or necessary without being necessarily necessary. By distinguishing carefully between different ways of formulating supervenience theses that are equivalent only on the assumption of the S5 axiom, and being selective about which he endorses and which he rejects, Wedgwood can avoid the direct arguments from supervenience against non-reductive moral realism. And because something can be necessary without being necessarily necessary, Wedgwood can explain the actual necessary connections between the normative and the non-normative by appeal to contingent facts—for though there are some such necessary connections, it is only contingent which ones there are. Schmitt and I try both to illustrate how resourceful this strategy is, and to show how many of Wedgwood's other commitments are needed in order to make it plausible. In particular, we argue that Wedgwood cannot plausibly make the distinctions that he needs among which formulations of supervenience are true and which are not without appealing to his apparently independent view of the interdependence of the normative and the mental.

But if we don't accept Wedgwood's strategy for responding to these modal challenges to non-reductive moral realism, there may still be hope for the non-reductivist. In *Being Realistic About Reasons*, Scanlon offers a response to such worries that is based on a thought I find illuminated by consideration of Richard Price's *A Review of the Principle Questions in Morals*. 'The Price of Supervenience' explores this strategy and its commitments, and argues that on its most promising version, it is driven by—you guessed it—the Standard Model Theory about normative explanations.

One of the most striking commitments of this strategy, I argue, is that in order to work, it must turn out that the basic normative concepts are monadic, rather than relations. Though this appears to conflict with the popular idea that reasons are the basic normative concept, we already saw in 'Reasons and Agent-Neutrality' how an advocate of Standard Model explanations of reasons could reduce the three-place relation *R is a reason for X to do A* to the dyadic relation, *R is a reason to do A*. But recall that one of my chief arguments against this view turned on the fact that when we come to basic reasons, it is hard to say what the reason is, to do these things. One response to that argument is to claim that in the basic cases, there need be no reasons that do the supporting of the reason-supported actions—there is only the monadic property of actions, of being reason-supported (which does not mean that there are reasons that support it). That this move both addresses my arguments against this view, and is also just what is

needed in order to use this view to help address the modal challenges to non-reductive moral realism, is I think strong evidence of the cohesive power of the Standard Model Theory.

4 Instrumental Rationality

The traditional testing-ground for the contrast between reasons that are reasons of only some, and reasons that are shared by necessarily everyone, is the case of instrumental rationality. It is widely agreed that there is some important phenomenon of means–end or instrumental rationality, and it is an intelligible and important view that this is all there is to rationality. One sort of critic—the Kantian critic—of this instrumentalist view holds that it goes wrong by failing to appreciate that there is more to rationality than instrumental rationality. But a deeper, rationalist, line of criticism traceable to Standard Model-like ideas, claims that the instrumentalist view is not just false, but incoherent, because instrumental rationality *must* itself be derived from the very same kind of objective, shared requirement of which the instrumentalist is skeptical. A third kind of critic takes a middle way: even if the instrumentalist perspective is not incoherent, the *best* theory of instrumental rationality, in its own right, reveals it to be governed by such an objective, shared, requirement. The essays in Part 3, 'The Scope of Instrumental Reason' and 'Means–End Coherence, Stringency, and Subjective Reasons', take up the issues involved in evaluating this third line of criticism of the instrumentalist.

If the instrumentalist has a thought that instrumental rationality is different in kind from other rational demands—as opposed merely to different in content—then it must be based in the idea that instrumental rationality isn't really an objective, universally shared set of demands, but rather requires different things of different people, depending on their ends. This perspective is challenged by the idea, particularly popular over the last few decades and famously championed by John Broome, that what is really required by instrumental rationality is simply that each agent satisfy the condition of intending the means she believes to be necessary to her intended ends. This is often called the 'wide scope' theory of instrumental rationality, because it appeals to a requirement to satisfy a conditional, as opposed to a requirement that is only conditional. (Thus, the requirement takes 'wide' scope over the conditional.) This requirement doesn't vary from agent to agent, and so it really is an objective, universally shared demand of rationality. And clearly if this is the best theory of instrumental rationality—even if it is not the only one—then if we wish to countenance instrumental rationality, we should make ourselves comfortable with the idea of such objective, universally shared demands. And as the pluralist strand in Standard Model theorizing always reminds us, if there can be one such demand, why not others?

'The Scope of Instrumental Reason' takes up the question of the scope of the requirements of instrumental rationality with precisely this dialectic in mind. It was my first

published paper, and I no longer agree with all of the ways in which I framed the issues in the paper, and in many places wish that I had taken greater care about how I finessed certain issues. But for all that, I still think that the basic ideas lying behind the 'symmetry' and 'agent-neutrality' objections to wide-scope theories in each domain are basically right.

'Means–End Coherence, Stringency, and Subjective Reasons' offers a later take on the topic of instrumental rationality, more carefully informed by consideration of the topic of instrumental rationality in its own right, and by the important differences between intentions and desires. I no longer believe that the actual account that I defend in this paper is on the right track, but I think that all of the arguments that I give about what we should expect a proper account to look like are still on track. My current view is that Jonathan Way's elegant papers, Way [2010] and [2012], develop and present the most promising contemporary theory of instrumental rationality that I know—one which takes advantage of everything that I argue to be important, in this paper, but relies on much more plausible and compelling assumptions than my own positive view. Though he would resist this conclusion himself, I even think that Way's view receives powerful intuitive support from symmetry considerations of the kind that I pose against other wide-scope theories.

In short, I don't think that the value of the two papers included in this part of the volume consists in that they offer theories that will be strong contenders for what we can ultimately accept. But I do think that they help to connect the challenges of explaining versus appealing to shared but unexplained obligations to distinctive normative challenges in offering the most adequate descriptive account of the domain. The 'agent-neutrality' objection to wide-scope theories is concerned with the former problem—of what explains these shared obligations, or where they come from. And the 'symmetry' objection is concerned directly with their descriptive adequacy. If I am right about this relationship, then there may be an intimate connection between what we try to explain and how we try to explain it, and the ultimate defensibility in their own right of the normative theories that we go on to give.

5 Autonomy

Because 'wide scope' theories of instrumental rationality have been popular over the last few decades, it is not surprising that charitable readers have sought to find this view in Kant's own account of hypothetical imperatives. The most famous defense of this interpretation of Kant is Thomas Hill's 1973 paper, 'The Hypothetical Imperative', but it is also defended by Korsgaard, Rawls, and others. Yet given what we already know in this introduction, this should be a very surprising interpretive claim about Kant. For we know that Kant thought that the possibility of hypothetical imperatives is philosophically unpuzzling, requiring only an analytic argument. And we know that Kant held the possibility of a categorical imperative to require a synthetic argument.

Yet what makes an imperative categorical is that it is unconditioned by any end—and according to the wide-scope theory, instrumental rationality is backed up by an unconditional requirement, binding on any rational agent as such, requiring her to conform her beliefs and intentions to the condition of means–end coherence. Kant's attitude toward the relative philosophical problematicness of hypothetical and categorical imperatives therefore makes him a very unlikely bedfellow of those who wish to explain the apparently conditional hypothetical imperatives in terms of an unconditional requirement of rationality.

In 'The Hypothetical Imperative?' I develop this point, and argue in greater detail on textual grounds against the wide-scope interpretation of Kant. This interpretation fails to make sense not only of Kant's claims about the distinctiveness of categorical imperatives, but of a distinction that he makes between 'problematic' and 'assertoric' hypothetical imperatives, and of his analytic argument for the possibility of hypothetical imperatives. I take it to be important where Kant stood on this question not only because the issues involved deeply inform our understanding of such basic issues in Kant's theory as the nature of imperatives and the argumentative structure of the *Groundwork*, but because if we misunderstand Kant on this point, then we will miss the truly deep and powerful insight that he offers in the course of his explanation of why categorical imperatives require a synthetic argument:

> By contrast, 'How is the imperative of morality possible?' is beyond all doubt the one question in need of a solution. For the moral imperative is in no way hypothetical, and consequently the objective necessity, which it affirms, cannot be supported by any presupposition, as was the case with hypothetical imperatives. [Kant 2002, 220 (4:419)]

According to Kant, it is precisely because hypothetical imperatives are conditional, that it is easy to explain them. We explain them by appeal to their condition. When it comes to unconditional imperatives, in contrast, there is no such condition to appeal to, in order to explain them.

This, I think, is a big, fundamental idea, and it is no coincidence that this quotation recurs repeatedly in essays throughout this volume. This idea links the structure of requirements to the possibilities for explaining those requirements. And the point is not specific to imperatives—it can be made for reasons, obligations, duties, or any of a variety of other normative concepts. It tells us that if we aspire for our normative theorizing to be explanatory and to find deep and general explanations, then we need to take great care over the structure of the phenomena that are to be explained.

In 'Hypothetical Imperatives, Scope, and Jurisdiction', I return to the issues raised in 'The Hypothetical Imperative?' in light of developments in the theory of instrumental rationality over the last ten years. I re-frame the terms of the scope debate, and argue that in order for the debate to make sense, we need to think in terms of a normative concept that is law-like, in that it admits of the concept of a *jurisdiction*. Laws don't just require things, they require things *of* particular people—the people who fall under their jurisdiction. What is at stake in the scope debate, I argue, is ultimately whether

the best explanation of the fact that someone who fails to intend the means she believes to be necessary to an end she intends is guaranteed to be irrational, consists in the fact that there is some norm whose jurisdiction includes every rational agent, and which proscribes such a condition for every such agent.

I argue that it does not. It is puzzling, I argue, where such rules come from and how they get to have jurisdiction over every rational agent, no matter what she is like. In contrast, I argue, it is less puzzling how each agent gets to have authority over herself. If the general phenomenon of means–end coherence is just a reflection of how we as agents require things of ourselves, then we can avoid the strange puzzles of what this heteronomous source of rational requirements is, and of how it comes to have authority over all rational agents, by taking advantage of Kant's insight that if a requirement is conditional, then we can explain it by appeal to its condition. Because the conditions of conditional requirements of rationality are psychological, we can take them to be exercises of a power each agent has to impose obligations on herself—a power that is literally auto-nomy, the power to legislate requirements for oneself.

The final essay of the volume, 'Scope for Rational Autonomy', continues to develop this thought, and pursues it from a slightly different angle. In his 2007 paper 'Requirements', John Broome gives a characterization of what is at stake between wide and narrow scope theories of rational requirements that closely matches my characterization in 'Hypothetical Imperatives, Scope, and Jurisdiction', but goes on to give two new arguments in favor of the wide-scope view over the narrow-scope view. This paper responds to Broome by offering a diagnosis of the source of the apparent force of his two arguments. Their force comes, I suggest, from a picture on which rationality, by being the 'source' of rational requirements, is analogous in some way to a legislator—the very kind of heteronomous legislator whose authority I argued to be suspicious in 'Hypothetical Imperatives, Scope, and Jurisdiction'. By presenting only a static model for rational requirements, Broome subsumes both wide- and narrow-scope views to this heteronomous picture.

To solve this problem, I present a fuller model which allows us to think at the same time about what the laws are, and about how those laws are legislated. My model allows us to distinguish between two importantly different narrow-scope theories of rationality—one of which we can agree is problematic in the ways pointed out by Broome. But it is the other of these views—one which is built on the idea that the rules of rationality are rules that we impose on ourselves—which makes good on the promise of narrow-scope theories to solve the explanatory problems that I have claimed best motivate narrow-scope theories all along. This view understands rational agents to be autonomous, self-legislating agents. And on this view, apparent shared obligations to be means–end coherent or to have consistent beliefs or intentions are really just reflections of facts about *how* we impose these rules on ourselves—not deeper, more general rules some external source.

Thinking about the rules of rationality as autonomous rules is therefore, I think, a fruitful perspective which brings together the importance of Kant's insight that

conditions give us tools with which to explain, and uses it to shed light on several puzzles raised by the idea that the rules of rationality have some external source. But it also brings us back to our fundamental worries about the explanatory burdens of shared reasons or obligations, and offers us a different and fruitful model for explaining them. One way, of course, in which it can turn out that there is a reason shared by every rational agent, is if each rational agent, merely in virtue of being a rational agent, is guaranteed to satisfy whatever it takes to have that particular reason. This is the strategy that I employed in *Slaves of the Passions*, and which comes up in the essays in Part 1.

In contrast, the idea explored in Part 4 is that some cases in which it appears that there is a shared reason or obligation are really just reflections of the *conditions* under which different individuals come to have the reasons or obligations that they do. In general, it is true of everyone that if they believe both p and $\sim p$, then they are failing to comply with some rule of rationality. But it doesn't follow that there is any rule of rationality that every such person is failing to comply with. The autonomous picture of where the rules of rationality come from advocated in the essays in Part 4 explains why not.

It has sometimes been thought—call this the traditional idea—that the claims of rationality are less philosophically problematic than the claims of morality, in part because rationality requires only 'consistency' of us, while morality actually requires something substantive. Philosophers like Jean Hampton have pushed back against this idea, on the grounds that even consistency is a substantive thing to be required, and Korsgaard has argued forcefully that only a chauvinist perspective that is already prejudiced against the requirements of morality can lead one to see any relevant difference between which of these requirements is more philosophically problematic. But if we put the ideas of Part 4 together with my arguments in *Slaves of the Passions*, we get the outlines of an argument that the traditional idea is right after all.

For the autonomous strategy outlined in the last few papers in this volume offers a powerful way to explain away the appearance of universal requirements of rationality, as instead merely the reflection of the way in which agents impose requirements on themselves. This strategy is made possible by the fact that what is apparently universally required is a kind of consistency between different states—only on such a view can we see the agent as being the source, through her adoption of one of these states, of a requirement not to adopt each of the others. But since the central requirements of morality do not appear to involve any such consistency, the only way that they could be given a similar treatment is on the basis of a philosophical argument for a shared relationship, that all agents necessarily have in common, to the action in question. This, of course, was the project I was engaged in in *Slaves of the Passions*.

6 Looking Backwards, Looking Forwards

I'll close by noting that most of the papers in this volume draw extensively on selective engagement with lessons from the history of moral philosophy. Indeed, I don't see how

they could do otherwise. The broader the lessons about the nature of and constraints on explanatory normative theorizing that I wish to draw, the more surprising it would be, if these lessons were not written all over the works of great figures of the canon who have been most engaged with exploring the shape and limits on moral explanation: Kant, Sidgwick, Moore, Ross, Mill, Price, Clarke—and Cudworth, just to name those who feature most prominently in my discussion.

In some cases I have found myself engaging in detailed textual analysis of how to interpret these figures—particularly Kant and Cudworth. Although I find these questions of textual analysis interesting in their own right, the passages I have been most concerned to engage with in this way are in each case some of the most widely quoted and in my view, least understood, passages from Cudworth and Kant. I have endeavored to draw a picture on which Kant and Cudworth represent poles of a centuries-old debate about the limits and power of subsumptive normative explanations and about the corresponding need to explain shared obligations. And in keeping with the progression of essays from Cudworth-focused to Kant-focused, I've taken Kant's side against Cudworth. If I'm right, then though Kant may not grasp and certainly would not endorse the whole picture of normative explanation advocated in this book, his basic conception of the explanatory project in moral philosophy still latched onto some of the most important insights. And that is as much as any philosopher could reasonably hope for.

Part 1

When an action, otherwise indifferent, becomes obligatory, by being made the subject of a *promise*; we are not to imagine, that our own will or breath alters the nature of things by making what is indifferent not so. But what was indifferent *before* the promise is still so; and it cannot be supposed, that, *after* the promise, it becomes obligatory, without a contradiction. All that the promise does, is, to alter the connexion of a particular effect; or to cause that to be an *instance* of right conduct which was not so before. There are no effects producible by us, which may not, in this manner, fall under different principles of morality; acquire connexions sometimes with happiness, and sometimes with misery; and thus stand in different relations to the eternal rules of duty.

[Price 1994, 51–52]

1

Cudworth and Normative Explanations

1.1 Understanding Normative Explanations

Moral theories do not purport merely to tell us which things we ought to do. They also try to tell us *why* we ought to do them. Moral theories, that is, generally have explanatory ambitions. What they try to explain to us is not why we *think* we ought to do certain things, of course, or why we *do* some things, but why we *ought* to do things. Little has been said, however, in a general vein, about how moral, or more generally, normative, explanations *work*—what sort of thing they are, in what ways they are like and unlike explanations of non-normative or *descriptive* phenomena, and so on. And that is unfortunate—for given the importance of explanatory ambitions in moral theorizing, differences in expectations about how moral explanations can and can't work could potentially be playing an important role in underwriting disagreement about many other questions in moral theory. In fact, I think that this is the case. But the only way to see whether this is so is to look hard for implicit theories about how moral or normative explanations must work.

The best way to look for implicit theories about normative explanations is to look for arguments which, once spelled out carefully, turn out to *need* assumptions about such explanations in order to work. In this paper I want to closely examine such an argument. It is originally due to Ralph Cudworth, and it is one of at least four different arguments that he offered against *voluntaristic* ethical theories. Cudworth's argument has since been widely held to conclusively establish its result, it was very influential in the eighteenth century, and arguments like it have recently been reiterated or endorsed by philosophers like Jean Hampton and Christine Korsgaard.[1] Voluntaristic theories, as Cudworth understood them, say that obligations derive from *commands* or *decisions*. Those commands or decisions may be those of God (as with Ockham, Descartes, and the Calvinists), those of a temporal sovereign (as with Hobbes or Protagoras), or even those of anyone whatsoever (as Cudworth understood Epicurus to claim). Cudworth

[1] Hampton [1998], Korsgaard [1996], [1997a], [1997b].

held that his argument worked against all of these views, and following Cudworth, Richard Price held that it worked against many other views as well.[2] But for concreteness, it's easier to focus on a single view.

1.2 Cudworth and Voluntarism

So consider *theological* voluntarism: the theory that every obligation derives from one of God's commands.

Voluntarism: For any person x and action-type a, if x ought to do a, that is because God has commanded x to do a.

Cudworth argues like this:[3] the voluntarist has to admit that in order for his theory to be true, God—or at least His commands—have to be pretty special. After all, we can all agree that when I command you to do something, it doesn't become the case that you ought to do it. So God's commands have to be different in some way from mine—they have to have *authority*, as Cudworth puts it. But what does the authority of God's commands consist in? Surely just this: that you *ought to do* what God commands. It is surely because you *ought to do* what God commands, while it is not the case that you ought to do what I command, that when God commands you to love your neighbor as yourself, you ought to do that, while when I command you to bring me my slippers, it has no such effect. And surely even the voluntarist has to agree with that much:

Authority Vol: $\Box\forall x$ (x ought to do what God commands)

But that is exactly what we need to get the voluntarist into trouble. For according to voluntarism, *every* time that you ought to do something, it is because God has commanded it. But why ought you to do what God commands? According to the theory, this would have to be because God has commanded it. But that is surely incoherent.

[2] Cudworth's *A Treatise Concerning Eternal and Immutable Morality* is the original source for the argument [1996, 17–22]; it is prominently picked up by Richard Price in his *A Review of the Principal Questions in Morals* [1994, 50–56]. There is a case to be made that Samuel Clarke ran the argument in his second Boyle lectures of 1705, published in 1706 as *A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation* [1967, 221–222], and as we'll see later, Christine Korsgaard cites Clarke for the argument. Clarke's lectures were published before Cudworth's *Treatise* was, even though they were a couple of decades after Cudworth would have written, so he may have some claim to independence. But the passage from Clarke is much less clear, and I think it is controversial whether he is really offering the same argument, so I'll focus on Cudworth and Price in what follows.

[3] "Wherefore since the thing willed in all laws is not that men should be bound or obliged to obey, this thing cannot be the product of the mere will of the commander, but it must proceed from something else, namely the right or authority of the commander [. . .] and an *antecedent obligation* to obedience in the subjects. Which things are not made by laws, but presupposed before all laws to make them valid. And if it should be imagined that anyone should make a positive law to require that others should be obliged or bound to obey him, everyone would think such a law ridiculous and absurd. For [. . .] if they were not before obliged, then they could not be obliged by any positive law, because they were not previously bound to obey such a person's commands. So that *obligation to obey all positive laws* is older than all laws, and previous or antecedent to them." Cudworth [1996, 18–19], italics added.

God could not make it the case that you ought to do what He commands simply by commanding it—if it weren't already the case that you ought to do what He commands, then such a command would make no difference, and if it were already the case that you ought to do what He commands, then it would be beside the point. So it follows that voluntarism is incoherent.

This *Cudworthy* argument—for as I'll be arguing, it is well worth chewing on again—obviously trades on the fact that the voluntarist's view is one about what *explains* why we ought to do things. If the voluntarist only believed that you ought to do something *just in case* God has commanded you to do it, or even *necessarily* just in case He has commanded it, then the Cudworthy argument would get no grip. For it might very well be that God *has* commanded you to do what he commands—little good though such a command does. Cudworth's claim is simply that a command like this can't be what *explains* why you ought to do what God commands. For according to Cudworth, that would already have to be true, in order to have such an effect. This is what makes Cudworth's argument an initially good candidate for illustrating something of how philosophers think about *explanations* of normative phenomena—such as what you ought to do.

1.3 The Argument Generalizes

Theological voluntarism is less popular these days than it once was, at least in philosophical circles. And this is largely because of Cudworth's argument, which was also widely promulgated in the eighteenth century by Richard Price.[4] So just in case you find the problem posed by Cudworth for theological voluntarism to be less than gripping, it is worth noting that a similar kind of argument can be run against *any* view that takes the form:

Theory: For all agents x and action-types a, whenever x ought to do a that is because x stands in relation $\mathfrak{R}$ to a.

Any perfectly general explanatory moral theory has the form of Theory. It is simply the form of a view that tries to give a perfectly unified answer to the question, "why ought

[4] Price [1994, 50–56]. Clarke is also often cited as having given the argument (see note 2), but I hesitate to attribute it to him. He certainly places much less importance on it; Price's intuitionist epistemology is driven by his metaphysical views, in favor of which he offers the Cudworthy argument—since moral truths are necessary ones, we need to apprehend them in the same way as we know other necessary truths [1994, 85]. But Clarke's metaphysical views are driven by his intuitionist epistemology—the *main* argument that he offers for the thesis that there are "unchangeable obligations of natural religion" is not the Cudworthy argument; he merely claims that his thesis is that "[t]hese things are so notoriously plain and self-evident, that nothing but the extremest stupidity of mind, corruption of manners, or perverseness of spirit, can possibly make any man entertain the least doubt concerning them." Clarke [1967, 194]. The passage in which Clarke *does* seem to be offering the Cudworthy argument is buried in his critical discussion of Hobbes, and it's not obvious to me that he doesn't mean, there, to be relying on the same appeal to intuition.

I to do *A*?" Yet a generalization of Cudworth's argument looks to rule out any such view entirely.

Consider: anyone who accepts Theory, after all, has to think that the relation $\mathfrak{R}$ is awfully special. Most relations you might bear to an action don't make it the case that you ought to do it, after all. You can be next to someone doing an action, or in a room where most people are doing it, or exactly thirty miles south of a piece of paper on which the action is written in Sanskrit, and none of these makes it the case that you ought to do it. So being related to it by $\mathfrak{R}$ must be special. It must have, as Cudworth would put it, *authority*. But what does this authority consist in, to explain why being related by $\mathfrak{R}$ to an action can obligate you to do it? It must be this: that you *ought to do* whatever action you are related to by $\mathfrak{R}$.

Now that—or so it seems—is exactly the sort of thing that Theory is supposed to explain—why you ought to do something. But if we *need* this, in order for the explanations offered by Theory to work, then it is hardly the sort of thing that Theory *could* explain. Imagine: if it were not already the case that being related by $\mathfrak{R}$ to an action obligated you to do it, then being related by $\mathfrak{R}$ to the action, doing-whatever-you-are-related-to-by-$\mathfrak{R}$, wouldn't make a difference. And if it were already the case, then it would not matter whether you were related by $\mathfrak{R}$ to it or not.

Now that was a mouthful. But on the face of it, it looks like a perfectly general application of Cudworth's argument. So if the Cudworthy argument successfully shows that not all obligations can be explained by God's commands, then it looks like it must also show that not all obligations can be explained by self-interest, by hypothetical contracts, by what would maximize the good, by what is in accordance with rules no one could reasonably reject, or *any* other source.[5] If this implication of the Cudworthy argument doesn't grip you, then I do not know what could. Cudworth's argument has long been held to *conclusively* establish the unviability of voluntaristic views. And in this section we've seen that all of the relevant steps in the argument look the same, against *any* perfectly general explanatory moral theory.

Whether you are convinced by Cudworth's argument or not, and whether you care about the viability of voluntarism or not, it would be *very* surprising if Cudworth's argument could show so much—*fascinating* if, like Cudworth and Price, you don't think that fully general explanatory moral theories like these are possible, and *threatening* if, like so many moral philosophers, you are at least occasionally tempted by the quest to find such a theory. Of course, like me, you may be one of those who find it obvious that, at least on the naïve literal reading that I've given it so far, Cudworth's argument has a gaping hole in it. And if so, then you should be puzzled at what could make so many moral philosophers find these arguments convincing in the first place, even in the face of their drastic consequences. The answer to this question is where I will look for a clue to how moral philosophers tend to think about normative explanations. It will lead us to a *theory* about such explanations that I think is initially attractive and commonly

[5] Although see section 1.12 for an important caveat we can make once we understand the argument better.

implicitly accepted, and this theory will allow us to develop a better, more *sophisticated* reading of Cudworth's argument.

1.4 A Questionable Inference

The problem, I think, once we look more closely, is not to discover whether or not the Cudworthy argument shows as much as it seemed to show in the last section, but to discover what makes it so tempting to think that it even got off of the ground in the first place. For in order to work, Cudworth's argument had to establish that even the voluntarist must agree that God's commands oblige you *because* you are under an *antecedent* obligation—to obey God. The problem for voluntarism is that its explanation of why you ought to love your neighbor is incomplete, until it posits a further obligation, and that though this obligation falls under the *scope* of the theory, it *cannot* be explained by the theory, because it is what makes the theory *work* at explaining everything else.

But why should we think that the voluntarist has to agree with this much? Here is what the voluntarist must agree with: she must agree that though God's commands oblige, mine do not. And she can agree that this is because God and I are different in an important way. And she can agree to stipulatively call this difference between me and God His *Authority*, and that He has *Authority* antecedently to any of His commands. But the problem arises when we try to say exactly what it is, in which God's Authority consists. As stipulated, it is simply whatever is necessary in order to explain the difference between me and God. What is in question, is whether the thing that is necessary in order to explain this difference is something that also falls under the scope of the voluntarist's explanatory theory.

Consider the way that we characterized God's Authority in section 1.2:

Authority Vol: $\Box\forall x$ (x ought to do what God commands)

But Authority Vol is *ambiguous*. It admits of two possible readings, depending on whether we read "do what God commands" transparently or opaquely.[6] On the transparent reading, Authority Vol tells us nothing more than that whatever God commands, we ought to do *that*:

Conditional Vol: $\Box\forall x\forall a$ (God has commanded x to do $a \rightarrow x$ ought to do a)

[6] Not everyone with whom I have discussed this paper has been able to generate both readings of Authority Vol. In particular, several have claimed not to be able to generate the opaque reading at all, taking it that there is clearly no such action as "doing what God commands" for the phrase "do what God commands" to refer to. But it should be obvious that I couldn't have succeeded in putting Cudworth's argument in the way that I did unless it was possible to get the opaque reading. Others, interestingly, have claimed not to be able to get the transparent reading of "do what God commands." This is also irrelevant to the main point. If there is no transparent reading of Authority Vol, then there is no sense in which Authority Vol uncontroversially sets out the claim that God has Authority, since we are understanding that claim to be whatever explains why God has, but I do not, the power to generate obligations by commanding others. All that is needed to explain this is Conditional Vol (see the main text).

According to Conditional Vol, when God commands you to do something, it becomes the case that you ought to do it. But Conditional Vol is not committed, on the face of it, to there being any action-type *A* such that you ought to do *A*. For example, *logically speaking* Conditional Vol could be true even though God has not commanded anyone to do anything. If that were so, then there *might* be *nothing* that you ought to do—for all that Conditional Vol says.

On the face of it, Conditional Vol is all that should be needed in order to explain the difference between me and God: that His commands, but not mine, generate obligations. Conditional Vol tells us that God has a certain status: that His commands lead to obligations. I lack that status. That seems to be precisely the difference between me and God that the voluntarist needed to explain by the Authority of God. So it looks like the argument that God has Authority could only commit the voluntarist to Conditional Vol, the transparent reading of Authority Vol.[7]

On the other hand, as I characterized the voluntarist, he aspired to explain why, when *X* ought to do *A*, this is so. He aspired to explain actual cases of when someone ought to do something. But Conditional Vol does not *logically* commit the voluntarist to any action that someone ought to do. So his account is not even supposed to apply to Conditional Vol. The account cannot, therefore, be circular by both being explained by, and seeking to explain, Conditional Vol. It does not even aspire to explain claims of the form of Conditional Vol.

On the other hand, there is another reading of Authority Vol on which it does commit the voluntarist to an action-type which someone ought to perform. This is the opaque reading:

Categorical Vol: $\Box\forall x$ (*x* stands in the *ought to* relation to the action-type: *doing whatever God commands*)

According to Categorical Vol, it is not merely the case that when God commands you to do something, it becomes the case that you ought to do it. It is already the case that there is something that you ought to do. You have "an antecedent obligation to obedience," as Cudworth himself puts it. There is *an action* to which you are already obliged.

[7] On the way I understand Arthur Prior's reading [Prior 1949, 13–25] (see note 10), all that we need in order to get Cudworth's argument to work is Conditional Vol. But this is enough to refute voluntarism only if we understand the voluntarist as holding that God creates not only all actual obligations, but all *conditional* truths about what obligations. Only that way would Conditional Vol fall under the explanatory scope of the theory. But no actual voluntarist is plausibly interpreted in this way. And Cudworth is quite explicit in claiming to prove that voluntarists must appeal to an "antecedent obligation," and not simply an antecedent conditional truth about obligations. So Prior's reading would make Cudworth's argument out be straightforwardly valid, but at the cost of depriving it of any real interest. J. A. Passmore's reading of the argument fits the one given so far, but he attributes to Cudworth a rather strong additional premise in order to make the argument work, similarly to what I will do in sections 1.5–1.7, but with less motivation [1951, 46]. (His discussion, moreover, is even more compressed than Cudworth's.) Stephen Darwall also discusses Cudworth's argument very briefly [1995, 118], but I'll come back to his reading in note 20. Interestingly, J. B. Schneewind doesn't mention the Cudworthy argument at all in any of his discussion of Cudworth, Clarke, and Price in *The Invention of Autonomy* [1998, 205–210, 310–323, 380–388].

Now, the voluntarist can accept Categorical Vol without incoherence. He may, for example, think that God *has* commanded (and would necessarily command) everyone to obey him. But he cannot accept without incoherence that Categorical Vol is the correct way of cashing out the thesis that God has Authority. For we stipulatively understood God's Authority as whatever is required to explain why, when God commands you to do something, it becomes the case that you ought to do it. So if Categorical Vol is the right way of understanding the claim that God has Authority, then the voluntarist's *complete* explanation of why you ought to be nice to your neighbors goes like this: God has commanded you to be nice to your neighbors, and obeying God is something that you ought to do. But if this is the form of explanation provided by the voluntarist, then it cannot be applied to explaining why you ought to do what God commands, in the sense of Categorical Vol. But Categorical Vol is precisely the kind of claim that the voluntarist aspired to explain. So if we assume that Categorical Vol is the way that the voluntarist has to understand the claim that God has Authority, then Cudworth's argument *works*.

The problem is this: Cudworth's argument *can* succeed at committing the voluntarist to appealing to Conditional Vol in order to explain the difference between me and God. But it *needs* to commit him to appealing to Categorical Vol. And the problem is that Categorical Vol simply does not follow from Conditional Vol, except by blatant equivocation. This makes it look like the Cudworthy argument simply turns on this blatant equivocation. That is the obvious hole in Cudworth's argument, on the naïve reading that we've given it so far.

1.5 An Explanatory Inference

But I think that we can do better for Cudworth. For one thing, in my experience, most philosophers who find Cudworth's argument gripping are not at all put off when the difference between Conditional Vol and Categorical Vol is pointed out to them. So it would at least be worthwhile arriving at an explanation for this *sociological* phenomenon about moral philosophers. And I hold that we can do better than a sociological explanation. I'll now show how to develop a *sophisticated* reading of Cudworth's argument—one that can really show that the voluntarist must appeal to Categorical Vol in order to complete the explanation of why God's commands can oblige you.

The thing to notice is this. For Cudworth's argument to work, it would not be enough if Categorical Vol was merely a logical consequence Conditional Vol. For the argument to work, Categorical Vol has to be needed to *complete* the explanation of why being commanded to by God makes it the case that you ought to love your neighbor. But so far as we've seen, all that the argument establishes is that Conditional Vol is needed in order to complete this explanation. If Categorical Vol *were* a logical consequence of Conditional Vol, that would not be enough to make *it* part of this explanation. So as long as Categorical Vol and Conditional Vol are *distinct*, Cudworth's argument needs something different from the thesis that Categorical Vol *follows* from Conditional Vol—which is convenient, because it doesn't.

In order to work, Cudworth's argument needs the thesis that Categorical Vol is *needed* in order to *explain* Conditional Vol, and thus to *complete* the explanation of why you ought to be nice to your neighbors. What the argument needs, in short, is a thesis about how explanations of normative phenomena have to work. And it is in search of this thesis that we've been investigating Cudworth's argument all along.

1.6 A *Grounding* Obligation

The pressure behind the idea that a thesis like Conditional Vol must be explained by one like Categorical Vol is not, in fact, hard to diagnose, and in this section we can draw it out in two steps, by comparison to two explanations—one non-normative and one normative—which seem like they *do* and *have* to work in this way. Take the non-normative explanation first. This is the explanation:

Anchorage: Anchorage is accessible by car from Philadelphia because Destruction Bay is accessible by car from Philadelphia.

By the same reasoning that was supposed to commit the voluntarist to Conditional Vol, this would not be an *adequate* or *complete* explanation unless the following were true:[8]

Conditional Anch: If Destruction Bay is accessible by car from Philadelphia, then Anchorage is accessible by car from Philadelphia.

But no one thinks that this is where the explanation ends. On the contrary, Conditional Anch cannot float free, without anchorage in the fact that:

Categorical Anch: Anchorage is accessible by car from Destruction Bay.

It is *obvious* that Conditional Anch is true only if and because Categorical Anch is. The thought about normative explanations that I want to pursue starts with the idea that Conditional Vol is like Conditional Anch—it needs to be explained or *grounded* by something else, and the thing it must be grounded in has to be a property of the *same* kind that is being explained. So, since what is being explained is why Anchorage is accessible by car from somewhere, what needs to be appealed to is that Anchorage is accessible by car from somewhere else. Similarly, in the Cudworthy argument, since we are explaining why some person ought to do some thing, what we have to appeal to is that she ought to do some other thing: to *obey God*.

Whether this is how things work in the Cudworthy argument should, I take it, be controversial. But it is easy, in fact, to isolate cases in which things *do* seem to uncontroversially look like they work similarly to the Anchorage case. So let's focus on the case of Rachel. Rachel is a liberal-arts undergraduate enrolled in a poetry-writing class, in which Professor Smith gives her the following assignment: "every morning

[8] It might be controversial whether an "adequate" explanation needs to appeal to any such conditional. But if it doesn't, then the voluntarist isn't even commited to Conditional Vol, and the argument doesn't even get off the ground.

when you get up, spend an hour writing about what you are thinking about." Since that is what was assigned, I take it that we can uncontroversially allow that here is an action that Rachel ought, on each morning, to do: to write about what she is thinking about.

Categorical Rachel: Every morning Rachel ought to do this: write about what she is thinking about.

But of course, doing what she ought to do every day is going to lead Rachel to act *differently*, on each day. So when she wakes up late on Monday with thoughts of regret for her feather pillows and dread of snoring through poetry class, writing about what she is thinking about is going to have to involve writing about feather pillows. And when she wakes up early on Thursday full of anticipation of her weekend plans, writing about what she is thinking about is going to have to involve writing about Phil's smile (for that is what lends excitement to her weekend plans). So on Monday not only is it the case that she ought to write about what she is thinking about, it is also the case that she ought to write about feather pillows. And similarly, on Thursday not only is it the case that she ought to write about what she is thinking about, it is also the case that she ought to write about Phil's smile. On each day, that is, whatever she is thinking about, she ought to write about that.

Conditional Rachel: Every morning, for all x, if Rachel is thinking about x, then Rachel ought to write about x.

But there is no mystery about why Conditional Rachel is true. Since we have already settled that it is uncontroversial that Categorical Rachel is true, it is clear that that is what explains why Conditional Rachel is true. It works this way: the soporific Professor Smith's assignment grounds Categorical Rachel, Categorical Rachel grounds Conditional Rachel, and Conditional Rachel, together with the facts about what Rachel is thinking about on each day, grounds the facts that on Monday she ought to write about feather pillows, on Thursday she ought to write about Phil's smile, and so on.

Rachel gives us a clear case in which a normative explanation works in the way that they have to *all* work, in order for Cudworth's argument to have the force that it requires. If all normative explanations have to work this way, then the explanation of Conditional Vol will have to work in this way. And if it does, then it must be explained by Categorical Vol. And that is what we need to make the Cudworthy argument work. It is a thesis about how normative explanations *have* to work.

1.7 The Standard Model for Normative Explanations

The explanation of why on Thursday Rachel ought to write about Phil's smile had three important features worth keeping track of. When we say, "On Thursday Rachel ought

to write about Phil's smile because that is what she is thinking about," what we have given is only a sketch of the complete explanation. For the explanation to be *complete*, it must (1) appeal to some further action that Rachel ought to do on Thursday, which (2) she ought to do whether or not she is thinking about Phil's smile, and such that (3) the fact that on Thursday she *is* thinking about Phil's smile explains why writing about Phil's smile is a *way* or *means*, in some broad sense, for her to do this other thing. In Rachel's case, the *further* action that she *antecedently* ought to do is to write about what she is thinking about. The fact that she is thinking about Phil's smile is *why* writing about Phil's smile is a way to do *this*, and *that* is why it is part of why she ought to write about Phil's smile—since she already, independently, ought to write about what she is thinking about.

Since it is *typical* or *standard* to expect explanations of why someone ought to do something to work in this way, I call this *model* for how normative explanations can work the *Standard Model*. To be more precise, an explanation follows the Standard Model insofar as it has the three features catalogued in the previous paragraph:

Standard Model: The explanation that *X* ought to do *A* because *P* follows the *Standard Model* just in case it works because there is (1) some further action *B* such that *X* ought to do *B* and (2) not just because *P* and (3) *P* explains why doing *A* is a *way* for *X* to do *B*.

And correlatively, the view that *all* normative explanations have to work in this way, we can call the *Standard Model Theory*:

SMT: For all *x*, *a*, and *p*, if *x* ought to do *a* because *p*, that explanation must follow the Standard Model.

The Standard Model Theory seems to be what we need in order to get the Cudworthy argument to work. The voluntarist believes that whenever you ought to do something, that is because God has commanded it. For this explanation to follow the Standard Model, it must appeal to some further thing that you ought to do, and not just because God has commanded this thing. And since that further thing that you ought to do also falls under the scope of the theory, the explanation of why you ought to do it must appeal to the same thing—namely, itself. But that makes the explanation circular, which is what gets us Cudworth's conclusion. So that is my provisional diagnosis of the force of Cudworth's argument: its key premise is a theory about how normative explanations *have* to work: the Standard Model Theory. This gives us a *sophisticated* reading of Cudworth's argument.

In support of this diagnosis, it is worth pointing out that the early proponents of the Cudworthy argument, including Cudworth and Price, were some of the only philosophers to explicitly articulate the Standard Model. Cudworth's moral work, entitled *A Treatise Concerning Eternal and Immutable Morality*, whose title Price echoes approvingly, is not only committed to the thesis that the basic moral truths are eternal and immutable. It is actually committed to the thesis that *which actions* are right or wrong does not and cannot change from one possible situation to another: the things that you

ought to do are the *very same* as the things that I ought to do, and necessarily so, in the views of both Cudworth and Price, as well as Clarke.[9]

So how, then, could this be? Surely, for example, if I promise to meet Sally for lunch at noon at Plato's Diner, and you promise to meet April for lunch at noon at Ben's Chili Bowl, the right thing for me to do at noon is to show up at Plato's Diner, and the right thing for you to do at lunch is to show up at Ben's Chili Bowl—right? So surely there are some cases in which the right thing for me to do and the right thing for you to do are different. The British rationalists used the Standard Model Theory to explain this phenomenon away:

> No will, therefore, can render *any thing* good and obligatory, which was not so antecedently, and from eternity; or any action right, that is not so in itself [. . .] It is true, the doing of any indifferent thing may become obligatory, in consequence of a command from a being possessed of rightful authority over us: But it is obvious, that in this case, the command produces a change in the circumstances of the agent, **and that what, in consequence of it, becomes obligatory, is not the same with what *before* was indifferent**. [. . .] When an action, otherwise indifferent, becomes obligatory, by being made the subject of a *promise*; we are not to imagine, that our own will or breath alters the nature of things by making what is indifferent not so. But **what was indifferent before the promise is still so**; and it cannot be supposed, that, *after* the promise, it becomes obligatory, without a contradiction. **All that the promise does, is, to alter the connexion of a particular effect; or to cause that to be an *instance* of right conduct which was not so before**. There are no effects producible by us, which may not, in this manner, fall under different principles of morality; acquire connexions sometimes with happiness, and sometimes with misery; and thus stand in different relations to the **eternal rules of duty**.[10]

According to Cudworth and Price, our promises change what the *ways* are for us to do the thing that we antecedently ought to do—to keep our promises—but that is only by changing which action-type showing up at Plato's Diner at noon is an *instance* of. It is not *qua* showing up for lunch at Plato's Diner that an action is *ever* right or wrong or

[9] Clarke's second Boyle lectures were entitled, *A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*. He says, "Some things are in their own nature good and reasonable and fit to be done [. . .] Other things are in their own nature absolutely evil" [1967, 196]. So Clarke doesn't merely believe that there are basic necessary truths (which might be conditionals) about obligations; he actually thinks that there are basic *actions* that are *fitting* or *unfitting*, and necessarily so. See also Price [1994, 51–52] and Cudworth [1996, 20–21].

[10] Price [1994, 52–53], boldface added for emphasis; italics in original. Cudworth also outlines the Standard Model:

> As for example, to keep faith and perform covenants is that which natural justice obligeth to absolutely. Therefore upon the supposition (*ex hypothesi*) that any one maketh a promise, which is a voluntary act of his own, to do something which he was not before obliged to by natural justice, upon the intervention of this voluntary act of his own, that indifferent thing promised **falling now under** something absolutely good and becoming the matter of promise and covenant, standeth for the present in a new relation to the rational nature of the promiser, and becometh for the time a thing which ought to be done by him, or which he is obliged to do.
>
> Cudworth [1996, 20], boldface added for emphasis; italics in original.

something I ought to do. It is only because it is an *instance* of keeping a promise, which is something that I independently ought to do, that I ought to do it. And that is precisely how Standard Model explanations work. So we really have cottoned on to what is doing the work in the Cudworthy argument, and it really is a distinctive and interesting theory about how normative explanations have to work.

1.8 Further Interesting Implications

In fact, I think, Cudworth's argument is only the beginning of the interesting implications of the Standard Model Theory. Understanding the Standard Model is crucial for appreciating the motivations and resourcefulness of historical ethical rationalist positions, and of importance to the particularism/generalism debate, answers to the question, "why should I be moral?" and much more.[11] This paper is not the place to become bogged down in all of the different problems in ethical theory in which I think that the Standard Model Theory can make an important difference. But let me catalogue just one more implication of the Standard Model Theory, just in order to make clear exactly how strong a theory it is.

It is a corollary of the Standard Model Theory that if normative differences between what different individuals ought to do can be explained at all, they must be explained by things that *everyone* ought to do. For example, take anything that some people ought to do, but not others. If this difference can be explained, the Standard Model Theory tells us that it must be explained by something that each ought to do. If there are yet others who ought not to do this, then *that* difference can only be explained if there is yet something further, which all ought to do. So either some differences must

Prior notes that "[t]o anyone acquainted with the Aristotelian logic, this use of the phrase 'falling under' immediately suggests the minor premiss of a syllogism" [1949, 20], recalling his contention that logic alone is enough to validate Cudworth's argument:

> We cannot infer 'We ought to do A' from, for example, 'God commands us to do X', unless this is supplemented by the ethical premiss, 'We ought to do what God commands'; and it is quite useless to offer instead some non-ethical premiss, such as 'God commands us to obey his commands'. [18–19]

Here the claim is that as a matter of *logic* we need to appeal to a further premise, and in the text Prior appeals to the *ambiguous* Authority Vol. If we are to agree with him that this is needed as a matter of logic, then we have to read it as Conditional Vol. And the same goes for Prior's example of how the Aristotelian syllogism would go: "To keep a promise is good; to do X is to keep a promise; Therefore, To do X is good" [20]. This *seems* to attribute *goodness* to the action of keeping promises, and so it *seems* to follow the Standard Model. But if this is an Aristotelian syllogism, then it has the same form as, "men are mortal; Socrates is a man; therefore Socrates is mortal." And the major premise of *this* syllogism attributes mortality not to man as such, but to *individual men*. To get this reading of Prior's example, we would have to read his major premise in the *transparent* way: as saying that *those actions* that are ways of keeping promises are good, but not necessarily saying anything about the further action of keeping promises as such. So if we read Prior as making the strictly logical claim—as suggested by his title, *Logic and the Basis of Ethics*—then he is disagreeing with me about how to read what is going on here in Cudworth. But Prior's own example is naturally read in a way that fits the Standard Model.

[11] On the last of these counts, see Chapter 2, this volume.

be totally unexplained, or all differences are ultimately explained by something that everyone ought to do.

But this is a very striking view. For on the face of it, many forms of skepticism about the objectivity of morality are motivated by the idea that it is *easier* to understand how it could be the case that John, who likes to ski, ought to ski, and Jim, whose sister is Mary, ought to help Mary study for the lsat, than to understand how both John and Jim, no matter what they are like, ought to give their earnings to world hunger relief, or ought to love their neighbor, or ought not to kill people. These kinds of skepticism about the universality of morality are precisely skepticism that there are things that *everyone* ought to do, even though there are things that one or another individual ought to do. The problem about *morality*, is that *its* requirements are supposed to be things that *everyone* ought to do, no matter what they are like.

If the Standard Model Theory is right, this kind of pervasive concern about the objectivity of morality is not merely mistaken or over-hasty. It is incoherent. For according to the Standard Model Theory, it *could not* be the case that John ought to ski or that Jim ought to help Mary get into law school, unless there was something that everyone ought to do. So if the Standard Model Theory is right, whatever is puzzling about morality, it *can not* be that its demands apply to everyone, no matter what they are like. According to the Standard Model Theory, *everyone* has to acknowledge *some* such universal demands. The only question can be whether their *content* is moral or not.[12] Yet this kind of puzzlement is certainly widespread, and seems to be shared at least in principle by many philosophers who are ultimately not skeptical at all about the universality of morality. Even *Kant* held that categorical imperatives required *more* philosophical explanation than hypothetical imperatives, rather than less. And all of that is incoherent if the Standard Model Theory is true.[13] So here we have yet another important question for which the Standard Model is going to have important implications.

[12] "Motivational skepticism about practical reason depends on, and cannot be the basis for, skepticism about the possible content of rational requirements." Korsgaard [1986, 331–332]. Again: "Most philosophers think it is both uncontroversial and unproblematic that practical reason requires us to take the means to our ends. [. . .] The interesting question, almost everyone agrees, is whether practical reason requires anything *more* of us than this" [1997b, 215]. See Schroeder [2007a] for further discussion.

[13] Compare Jean Hampton:

> Kant's position on the nature of hypothetical imperatives **must** be construed (contra his explicit wishes) such that understanding the bindingness of a hypothetical imperative is no easier than understanding the bindingness of a categorical imperative. My interpretation cannot save Kant's belief that the former is more straightforward than the latter; indeed, my argument is that Kant's belief is wrong. The **only** way to analyze Kant's analyticity claim is to do so in a way that locates in hypothetical imperatives the same mysterious objective authority that attends the categorical imperative. Even more strikingly, I have argued that the force of hypothetical imperatives is dependent on, and *is at least in part constituted by*, the force of some antecedent categorical imperative that is in part definitive of instrumental rationality.
>
> Hampton [1998, 165–166], boldface added for emphasis.

1.9 Still a Puzzling View

The implications of the Standard Model Theory should be enough to demonstrate that it should be controversial—lots of interesting views in ethical theory must hold that it is false, on pain of incoherence. It has undeniable initial appeal which, as I've noted, consists in the idea that conditionals like Conditional Vol cannot be true all by themselves, but must somehow be grounded in categorical facts. This seemed particularly obvious in the Anchorage case. But as other examples can easily show us, there are *other* ways in which such conditionals can be grounded—*without* being grounded in the *kinds* of fact that Standard Model explanations require.

The problem is that the *accessible by car* case is actually quite unusual among non-normative explanations. *Very few* non-normative explanations work in a way that closely parallels the Standard Model. Most are more like the following case: Marcus is in Milwaukee. So he is north of Chicago. He is north of Chicago *because* he is in Milwaukee—if he left Milwaukee for St. Louis, for example, then he would no longer be north of Chicago. So consider how a Standard Model-like explanation of Marcus's situation would go. We can appeal to the following *conditional*:

Conditional Chi: If Marcus is in Milwaukee, then Marcus is north of Chicago.

For this to be a Standard Model-like explanation, then this would have to be true in virtue of someplace that Marcus is north of—the very kind of thing that we are trying to explain.

But there is no place that Marcus is north of that explains Conditional Chi. Even if Marcus goes to the South Pole, it will still be true that if he goes to Milwaukee, he will be north of Chicago—but he will not be north of anywhere. This *conditional* property of where Marcus can end up being north of is not grounded in a *categorical* property that *he* has, of already being north of somewhere. This is not to say that it is not grounded in *any* categorical property of Marcus, nor to say that it is not grounded in any categorical fact about what is north of what. On the contrary, it is grounded in the fact that *Milwaukee* is north of Chicago. All I am pointing out is that it is *not* grounded in a categorical fact about where *Marcus* is north of, as we would be bizarrely pushed to say if we sincerely tried to model this explanation on the Standard Model for normative explanations.

1.10 An Alternative to the Standard Model

I hope now to have established two broad theses: first, that the Standard Model Theory has a certain intuitive appeal, is widely accepted—at least implicitly—and plays a crucial role in whether Cudworth's argument successfully establishes the incoherence of theological voluntarism, let alone any of the wide range of views taking the form of Theory. And second, that the Standard Model Theory, once articulated, should be a *surprising* thesis, once compared to the Milwaukee case, and should in fact be highly

controversial even in ethics, given the kinds of consequences to which it leads. I now want to articulate a model for normative explanations that is opposed to the Standard Model—which leads to an alternative theory about how at least *some* explanations of normative phenomena can, at least *in principle*, work.

In order to see whether the Standard Model is the only way to explain conditionals such as Conditional Vol, it is worth looking at how we might explain *non*-normative conditionals like Conditional Chi. We did not need the analogue of the Standard Model in order to explain Conditional Chi. There doesn't have to be any place that everyone is north of, in order to explain how it could be true that everyone has the following property: if she goes to Milwaukee, she will be north of Chicago. So what explains the truth of Conditional Chi?

Conditional Chi, I think, is true because of two things. First, Milwaukee is a place north of Chicago. And second, for a person to be north of a place is just for her to occupy a *place* that is north of that place. That's why it is true that if anyone goes to Milwaukee, she will be north of Chicago. Going to Milwaukee is a way of occupying a place, Milwaukee, that is north of Chicago, and that is precisely what it takes to be north of Chicago. Other ways of being north of Chicago include being in any of Evanston, Racine, Sheboygan, or Green Bay. And that is because Evanston, Racine, Sheboygan, and Green Bay are all places that are north of Chicago.

This explanation relies on a constitutive truth: the fact that occupying a place that is north of a given place is *just what it is* to be north of the given place.[14] This constitutive truth explains[15] why it is true that:

Conditional Place: $\Box\forall x\forall p\,(\exists q\,(x \text{ is in } q \,\&\, q \text{ is north of } p) \rightarrow x \text{ is north of } p)$

Conditional Chi falls out from Conditional Place and the fact that Milwaukee is north of Chicago. Rachel's case gave us some reason to think that at least sometimes *normative* conditionals do not *have* to be explained in this constitutive way—they can be

[14] I won't pretend that this explanation is uncontroversial. It commits to substantivalism about space. A relationalist about space may think that Milwaukee counts as north of Chicago because *people* in Milwaukee are north of *people* in Chicago—in short, that the explanation goes the other way around. For my purposes, I don't care which of these is the case; I merely want to use the case to illustrate how a constitutive explanation might work. I could have used an uncontroversial example, about triangles or something, but this one was already salient in the discussion.

[15] Constitutive answers to questions are sometimes *contrasted* with explanatory answers. If, when asked why Zena is tired, I say that it is because of a certain chemical pattern in her bloodstream, I may have succeeded in giving a constitutive answer, but intuitively I haven't offered an explanation. Naturally there is a restricted sense of explanation in which we want something more specific. And in any normal context, explaining why a person is tired by noting what is going on in her bloodstream will not count as having answered the original question. But it does seem that "because Milwaukee is north of Chicago" is a perfectly good explanation of why if anyone goes to Milwaukee, they'll be north of Chicago. And I don't see any way to fill out this answer except by noting the constitutive truth that being north of a place *just is* being in a place that is north of that place. So there we do have a constitutive story that counts as giving an explanation. The constitutive explanation of Conditional Place is simply a limiting case. Unlike with the constitutive explanation of Conditional Chi, we don't have to mention anything at all other than the constitutive truths. The most basic conditional accepted by the voluntarist will be like that.

explained by appeal to a further, categorical, fact about what someone ought to do. We might say that such explanations follow a *different* model than Standard Model explanations, and we can call it the *Constitutive Model*.

1.11 Reduction and Constitutive Explanations

Suppose, then, that the voluntarist holds that just as *what it is* to be north of a place is to occupy a place that is north of it, *what it is* for it to be the case that someone ought to do something is just for God to have commanded her to do it:

Constitutive Vol: For God to have commanded *X* to do *A* is *just what it is* for it to be the case that *X* ought to do *A*.

Such a view would be a *reductive* view about *ought*s. It would analyze them in terms of something else—God's commands. But it would be an intelligible picture on which not all normative explanations follow the Standard Model. On this picture, though many normative explanations may follow the Standard Model, Standard Model explanations eventually run out, and when they do, the only further explanation of why it is the case that someone ought to do something, is to point to *what it is* for it to be the case that she ought to do it, and to point out that this, in fact, holds.

None of this move, however, hinges on the voluntarist accepting Constitutive Vol in particular. In fact, we may have some reason to suspect that a good voluntarist would not accept Constitutive Vol. After all, if you ask a voluntarist why we ought to do what God commands, you might expect him to say that it is because God is our creator, or our savior, or is all-loving. If this seems to him to be an appropriate answer, then he must not really accept Constitutive Vol. But he might still accept a *structurally similar* claim:

Constitutive Create: For *X*'s creator and savior to have commanded *X* to do *A* is *just what it is* for it to be the case that *X* ought to do *A*.

The voluntarist who accepts Constitutive Create can make all of the same moves against Cudworth's argument as the voluntarist who accepts Constitutive Vol; he just thinks that the constitutive answer lies farther along the line. Constitutive Create may be a more plausible view for other reasons, but in our dialectic, accepting it works in exactly the same way as accepting Constitutive Vol.[16] So for Cudworth's inference to be plausible, we need to have some reason to think that we *could not* replicate this style of constitutive explanation for Conditional Vol. But if it works elsewhere, why *couldn't* it work here?

[16] It is worth noting, on this score, that Robert Adams' version of Theological Voluntarism *is* a constitutive view. See Adams [1973], [1979]. Some authors, in fact, actually *define* Theological Voluntarism as such a reductive view. But compare Quinn [1979], [1990], [1999].

1.12 Additional Considerations

It may be that neither Constitutive Vol nor Constitutive Create seems to you to be a plausible view. Indeed, like many moral philosophers, you may believe that *no* view which says *what it is* for it to be the case that someone *ought* to do something in non-normative terms, supernatural or otherwise, could possibly be true. So my proposed *Constitutive* model for normative explanations, modeled on my explanation of the Milwaukee case, may not seem to you to be a particularly promising alternative to the Standard Model, and thus you may believe that it doesn't constitute a very good response on behalf of the voluntarist to the Cudworthy argument. But I think that even you should be able to agree that the Constitutive Model is, at least *in principle*, enough to let us off of the hook as far as the Cudworthy argument goes. For I think that the Constitutive Model can serve very well to explain one of the glaringly obvious features of how the Cudworth argument seems to work in relation to different explanatory moral theories.

As we saw in section 1.3, except for the questionable inference, the Cudworthy argument makes no discrimination between different explanatory moral theories. So it would seem to work just as well against Explanatory Consequentialism as against theological voluntarism:

Expl. Consequence: For all *x* and all *a*, if *x* ought to do *a*, that is because of the options available to *x*, doing *a* will bring about the most good.

A consequentialist doesn't have to accept Explanatory Consequentialism. Some consequentialists, perhaps for Standard Model reasons, think that there is one basic action that everyone ought to do at all times—to bring about the most good.[17] But it would be incoherent for Explanatory Consequentialism to think that this obligation was needed for its explanation, for it certainly can't explain itself.

One puzzle about the Cudworthy argument is why the argument is much more persuasive against voluntarism than against Explanatory Consequentialism. Many[18] who find it obvious that voluntarism is doomed because of Cudworth's argument are less worried about the prospects for Explanatory Consequentialism. I propose that if the Constitutive Model for normative explanations is the principal alternative to the Standard Model, that can serve to *explain* this reaction.

Suppose that some constitutive theories are more or less plausible or attractive than others. For example, Constitutive Vol might be much less attractive than

[17] According to James Dreier, consequentialism is committed to a "demand that each person seek to maximize the realization of what is of impersonal value" [1993, 22]. Shelly Kagan places great weight on the existence of a *pro tanto* reason to promote the good, to which he claims *everyone*—even non-consequentialists—are committed, and takes the difference between consequentialists and non-consequentialists to lie in whether there are any other, conflicting, *pro tanto* reasons [1989, 15–19].

[18] Though perhaps not all. Presumably Dreier and Kagan, if they are really sincere in holding that all consequentialists are committed to these basic requirements, rather than simply to a basic *conditional*, would have to think that Explanatory Consequentialism isn't a viable possibility.

Constitutive Create. And both might be considerably less plausible than Constitutive Consequentialism:

Constitutive Con: For it to be the case that *X* ought to do *A* is just for it to be the case that of the options available to *X*, doing *A* will bring about the most good.

There are a number of reasons why Constitutive Con might be more plausible than either Constitutive Vol or Constitutive Create. Chief among them might be that Constitutive Con employs another *normative* notion—the *good*, in order to explain what someone ought to do, rather than amounting to a reduction of a normative category to entirely non-normative categories. So though it is in one sense *reductive*, it does not seek to reduce any normative category in *non*-normative terms. But the important thing, for my explanation, is merely that it *is* a more plausible thesis.

If Cudworth's argument against voluntarism looks like a better argument than the Cudworthy argument against Explanatory Consequentialism, that has to have something to do with the Questionable Inference in Cudworth's argument. For otherwise the two arguments are exactly identical. On my interpretation, the Questionable Inference can be filled out by appealing to the need to *complete* the explanation of how God's commands can oblige, or how bringing about the most good can. The argument gets its force, then, from how much pressure we feel to follow the Standard Model in completing this explanation. So how compelling the argument seems should vary inversely with how plausible we find *alternative* modes of explanation. And if the Constitutive Model that I've sketched is the chief alternative to the Standard Model, then the difference in the plausibility of Constitutive Vol and Constitutive Con can successfully explain the difference in how compelling the Cudworthy argument seems, against each of these views. I take this explanation as some support in favor of my contention that my Constitutive Model for normative explanations is at least in principle one of the chief alternatives to the Standard Model.

But I can offer yet one more piece of historical evidence on this score. For having articulated the Constitutive Model as an alternative to the Standard Model, I think we can now see that on the best reading of Cudworth's original argument, the argument does *not*, in fact, rely on the full-fledged Standard Model Theory. For *both* Cudworth *and* Price, in running the Cudworthy argument, *take care to insist* that voluntarists can't really intend to mean that "ought to" just *means* "has been commanded by God to," and in favor of this point each offers an argument that anticipates Moore's Open Question Argument[19] (in fact, Price's argument is plausibly more careful than Moore's).

[19] Cudworth prefaces his discussion with the remark, "Wherefore in the first place, it is a thing which we shall very easily demonstrate, that moral good and evil, just and unjust, honest and dishonest (if they be not mere names without any signification, or names for nothing else but willed and commanded, but have a reality in respect of the persons obliged to do and avoid them) cannot possibly be arbitrary things, made by will without nature" [1996, 16]. Schneewind claims that this passage sets the reductive view aside without argument [1998, 208], but Darwall claims that this relies on a rudimentary argument that surely a voluntarist would not want to be committed to his thesis being a tautology, reminiscent of the Open Question argument [1995, 118]. In Price things are much clearer. He says, "Right and wrong when applied to actions which are

This makes the most sense if Cudworth and Price in fact did *not* accept the Standard Model Theory, but instead only something somewhat *weaker*: what we might call the *Standard-Constitutive Conjecture*:

SCC: All normative explanations must *either* follow the Standard Model *or* the Constitutive Model.

This gives us a *third* reading of Cudworth's argument. In section 1.2 I gave the *naïve* reading of the argument, which I showed in section 1.4 turned on an equivocation. The Standard Model Theory gave us a *strong* sophisticated reading of the argument in section 1.7. But now I can give a *weak* sophisticated reading of Cudworth and Price's argument. Instead of assuming the Standard Model Theory in all of its strength, they only need to assume the Standard-Constitutive Conjecture. This conjecture forces the voluntarist onto a dilemma. He must either accept a constitutive view or not. If he does, then it is supposed to fall to the Open Question argument. Whereas if he does not, then his explanation must follow the Standard Model, and the argument continues as on the *strong* sophisticated reading.[20] This is my considered interpretation of the historical Cudworth and the historical Price: both allow that there *could in principle* be non-Standard Model explanations, *if* there were any true reductive view about *ought*. Their arguments turned on having an *independent* argument against reductive views: the Open Question argument.

1.13 Cudworth and Reduction

But if that is right, then the Cudworthy argument has been misappropriated by more contemporary theorists. For one of the reasons why Cudworth's argument against voluntarism looks initially so interesting, is that it looks like it gives us an independent, perfectly general, argument that no *reductive* view about morality could possibly be true. For any reductive view about what we ought to do surely counts as attempting to give a unified explanation of every case in which someone ought to do something. So any reductive view is going to be committed to a thesis of the form of Theory. And the generalized version of Cudworth's argument purports to show that any view of the form of Theory is incoherent.

commanded or forbidden by the will of God, or that produce good or harm, do not signify merely, that such actions are commanded or forbidden, or that they are useful or hurtful [. . .] Were not this true, it would be palpably absurd in any case to ask, whether it is *right* to obey a command, or *wrong* to disobey it; and the propositions, *obeying a command is right*, or *producing happiness is right*, would be most trifling, as expressing no more than that obeying a command, is obeying a command, or producing happiness, is producing happiness" [1994, 16–17].

[20] On the weak sophisticated reading, then, Cudworth's argument has essentially the structure attributed to it by Stephen Darwall [1995, 118]. But Darwall's discussion of the second fork of the dilemma is too compressed for us to be able to tell whether he understands it in my way—what he says is compatible even with Prior's reading [1949, 18–19].

Christine Korsgaard, in fact, seems to endorse Cudworth's argument as decisively refuting any reductive view about morality. According to Korsgaard, there are only four ways of answering what she calls the "normative question," which is essentially the question, "why should I do what I ought to do?" But only one of Korsgaard's four answers, the first, *voluntarist*, answer, actually involves an answer of the form, "you ought to be moral because *p*." The second, *realist* answer, is no answer at all—it is just foot-stomping in insistence that *it is a fact* that you ought to be moral.[21] Korsgaard's third answer amounts to something like, "you already believe that you ought to be moral," and what her fourth answer really does is to explain why you *have* to believe that you ought to be moral.[22] So it is a kind of transcendental argument for the truth of "I ought to be moral," but it doesn't give us any answer of the form, "you ought to do what you ought to do because *p*." Only the first, *voluntarist*, answer does that. So if there is *any* place for a reductive view in Korsgaard's purportedly exhaustive classification of potential answers to the normative question, all reductive views must fall under the *voluntarist* category. And Korsgaard's argument against the voluntarist answer is just the Cudworthy argument, for which she cites Clarke.[23]

Yet if my diagnosis of what makes the argument compelling is right, then this is a misappropriation of Cudworth's argument. Korsgaard and, on one reading, Hampton,[24] would have us use the Cudworthy argument as an *independent* argument for the impossibility of reductive normative theories. If it *were* such an independent argument, that would help to explain why philosophers like Korsgaard and Hampton insist that reduction is thoroughly hopeless, even though it is widely acknowledged that the only generally discussed objection to reductive normative theories, the Open Question argument, fails to conclusively establish anything about *synthetic* reductive views.[25] The Cudworthy argument—at a first pass—looks like it *could* give us a deep

[21] "Having discovered that he needs an unconditional answer, the realist concludes straightaway the he has found one" [1996, 33].

[22] "Kant, like the realist, thinks we must show that particular actions are right and particular ends are good. Each impulse as it offers itself to the will must pass a kind of test for normativity before we can adopt it as a reason for action. But the test that it must pass is not the test of knowledge or truth. For Kant, like Hume and Williams, thinks that morality is grounded in human nature, and that moral properties are projections of human dispositions" [1996, 91]. See, in general, all of Korsgaard's lectures 2 and 3.

[23] "We can keep asking why: 'Why must I do what is right?'—'Because it is commanded by God'—'But why must I do what is commanded by God?'—and so on, in a way that apparently can go on forever" [1996, 33]. Also: "Samuel Clarke, the first defender of realism, was quick to spot what he took to be a fatal flaw in the view I have just described" [28]. See notes 2 and 4. Korsgaard also appeals to considerations that look like the Cudworthy argument elsewhere, for example, in explaining Kant's argument at [4:420–421] of the *Groundwork*, in her introduction to the Cambridge edition [1997a, xvii]. And it seems to be presupposed by how she sets things up in "The Normativity of Instrumental Reason" [1997b].

[24] Hampton [1998] (see note 13). In a nice but brief discussion, Richard Joyce infers that Hampton is committed to using the Cudworthy argument this way [2001, 115–123]. I tend to agree that this is the best way to make Hampton's arguments out to be interesting and have the kind of scope she intends for them. But Hampton's unfinished and posthumously published [1998] is *very* unclear, so it's not safe to stake too much on this interpretation—or on any other.

[25] Moore [1903, 61–68], Boyd [1988], Brink [1989, 144–170].

and general argument against reductive normative theories, and I think there is a good case that at least some philosophers have understood it in this way.

But what the Cudworthy argument shows, in fact, is merely that reductive views are committed to understanding some normative explanations as working in some non-Standard Model way. The stakes for demonstrating a view to be internally incoherent are high. If there is *any* way of understanding a view as not being internally incoherent, it is more charitable to understand it in that way. So at the very least, we have to understand a reductive theorist as *believing* that reductions can provide an alternative model for normative explanations to that provided by the Standard Model.

And I agreed in section 1.12 that non-reductivists should be able to agree with this, even though they think no reductive view is true. The fact that Cudworth's argument seems to show too much—even that Explanatory Consequentialism is incoherent—supports the view that some other model for normative explanations must be available. And the fact that reductions of some normative concepts to others are more plausible than reductions of the normative to the non-normative supports the view that what distinguishes Explanatory Consequentialism from voluntarism is precisely that Constitutive Con is more plausible than Constitutive Vol. So whatever you think of reductions of the normative, if you think that Explanatory Consequentialism stands up to the Cudworthy argument better than voluntarism does, then you ought to think that explanations of normative phenomena can at least *potentially* follow the Constitutive Model—*if* there are any true constitutive views.

But if they *can*, then Cudworth's argument shows a reductive view to be false only if we can assume that its constitutive explanation, like that of Constitutive Vol, can not work. And that means that Cudworth's argument does not show us *anything* about the general viability of reductive views. It shows them to be incoherent only by assuming that their reductions can not work. And so if *it* is going to provide a successful argument against the possibility of the reduction of the normative to the non-normative, then it is going to have to presuppose some *other, independent* general argument to the same effect. Cudworth and Price think that they have an independent argument for this conclusion, but that is only because they are convinced by the Open Question argument. If we are no longer convinced by the Open Question argument, then so far from being able to appeal to the Cudworthy argument in order to be able to bolster our case against reductive views, we are still in need of evidence against reduction in order to get the Cudworthy argument off of the ground. *Any* reductive view can offer a Constitutive Model explanation, and so any reductive view can evade Cudworth's argument.

1.14 Three Lessons

I now hope to have accomplished at least three things in this paper. First, I hope to have made sense of why it is that Cudworth's argument is so attractive to moral philosophers, despite the fact that it looks to trade on a blatant equivocation. The argument

is nevertheless attractive, because (1) conditionals like Conditional Vol, just like conditionals like Conditional Anch, seem to need to be grounded *somehow*, and (2) like Rachel's, so many other normative explanations *do* follow the Standard Model. So it is very *natural* for a moral philosopher to believe that all do.

Second, I think that this serves to help us answer a hard question in metaethics. Non-reductive realists about the moral or normative believe that reductions of the normative to the non-normative are obviously impossible, or that such reductive views are crazy.[26] The chief purported argument to this effect in the literature is the Open Question argument, but it is far from clear that it establishes that reductive views must be false, let alone that it is *crazy* to think that a sophisticated reductive view could be true. So why is it, then, that non-reductive theorists find reduction so hopelessly crazy? I think that there is more than one reason, and have written about some of the others elsewhere.[27] But the Cudworthy argument gives us one natural diagnosis. If you already find reductions of the normative to the non-normative implausible, you will think that the reductive theorist's constitutive explanation of why his theory works doesn't work.

By the Standard-Constitutive Conjecture, if it doesn't work, then you need a Standard Model explanation to explain why his theory works. And that has to appeal to something that falls under the scope of his theory, but can't, on pain of vicious circularity. So the reductive view looks hopeless. But if this line attracts you, you are ignoring the fact that the reductive theorist has different ideas about the possible range of models for explanations of normative phenomena. Because *he* thinks that his reductive view *is* true, he has no trouble accepting that it can successfully back up constitutive explanations of normative phenomena. If neither the non-reductivist nor the reductivist makes their theories about how their explanations are working *explicit*, it is easy to see how they could end up butting heads over this question.

And third, I hope to have succeeded in at least further articulating the problem with which I began: to characterize what assumptions are at work in moral philosophers'

[26] As Thomas Nagel puts his skepticism about reduction, "If values are objective, they must be so in their own right, and not through reducibility to some other kind of objective fact. They have to be objective *values*, not objective anything else" [1986, 138]. This expresses the idea that reduction is not a way of being realist about the normative at all. In fact, David McNaughton, in a book marketed as a textbook, explicitly categorizes reductive views as *irrealist*: "While such a reductive account ensures that moral views are true or false in virtue of facts that are independent of the speaker's opinion on the matter, it is nevertheless an irrealist position. For it does not allow that there are *distinctive* moral facts which are independent of our current opinions, waiting to be discovered by our moral inquiries" [1989, 44]. McNaughton maintains this despite the fact that the "distinctive" part didn't actually make it into his definition of moral realism [39]—he finds it so obvious as not to require further argument. According to Graham Oddie [2005, 18], it is obvious that reductive views are insufficiently realist because "[t]hat which is reducible is less real than that to which it reduces." And, of course, Derek Parfit has been known to say such things as that "it is *amazing* that these truths still need defending," when referring to such theories as the value-based theory of reasons and non-reductive normative realism (in an unpublished manuscript). Despite lack of clear, conclusive argument, many philosophers continue to think that it is *obvious* that reductive normative realism is false.

[27] Schroeder [2005b].

explanations of normative phenomena, and how those explanations are both like and unlike other kinds of explanation. On the Constitutive Model, normative explanations can sometimes work in ways that are very like explanations of non-normative phenomena, like the fact that if you go to Milwaukee you will be north of Chicago. But Standard Model normative explanations are much less like explanations of other kinds of thing—with implications, at least, for how we should think about the scope and tasks of explanatory moral theory, and the intelligibility of certain kinds of skepticism about the objectivity of morality.

And that only leaves us with more, and harder, questions than I can attempt to answer here. What other hard questions in moral philosophy are affected by how we think about how normative explanations have to work? How can we generalize and precisify the Standard Model to explanations of reasons, of what is good, and so on? What more can we say about how constitutive explanations work? Are there other good, independent, general reasons to think that such explanations aren't possible in the normative case? And most importantly, is the Standard-Constitutive Conjecture correct? Are there further, as yet unexplored, models for understanding how explanations of normative phenomena might work? Ones which commit neither to the implications of the Standard Model, nor to the reductive theses required in order to make constitutive explanations work? I wish I knew. The most that I can claim to have done is to have demonstrated that there is far more interesting work to be done, in articulating our implicit and sometimes surprising theories about how normative explanations are supposed to work.[28]

[28] Special thanks to Gideon Rosen, David Sussman, Stephen Darwall, Michael Morreau, Nic Southwood, David Sobel, Alasdair Norcross, Scott James, Mark Johnston, Gillian Russell, and Zena Hitz, to audiences at Northwestern University, Rice University, New York University, the University of Maryland at College Park, and Bowling Green State University, and to all the members of the Princeton graduate student dissertation seminar, for helpful or stimulating discussion of this or related topics. Thanks also, finally, to the brave students in Honors 218Z at the University of Maryland.

2

Reasons and Agent-Neutrality

2.1 Introduction

Consider the following sentences:

1 The fact that Katie needs help is a reason to help Katie.
2 There is a reason to help Katie.
3 The fact that there will be dancing at the party is a reason for Ronnie to go to the party.
4 There is a reason for Ronnie to go to the party.

Each sentence uses the word "reason" to express some relation, but the relation expressed by each is different. In **1** it expresses a *dyadic* relation between a consideration[1] and an action, in **2** it expresses a monadic property of an action, in **3** it expresses a *triadic* relation between a consideration, an agent, and an action, and in **4** it expresses a dyadic relation between an agent and an action. But it is natural, nevertheless, to think that these senses of "reason" are not unrelated.

It is natural, for example, to take the "there is" in **2** and **4** at face value, as an existential quantifier. After all, **2** is a consequence of **1** and **4** of **3**, in the same way that we would expect if it were. And if someone tells you that there is a reason to help Katie, it is perfectly fair game for you to ask, "well, what is it?" So the relationship between **1** and **2**, and between **3** and **4**, ought not to be particularly mysterious. I'm going to take the view that this is correct.[2] So in order to understand **2** we have to understand **1**, and in order to understand **4** we have to understand **3**. What I'll be interested in, in this paper, is therefore the relationship between **1** and **3**. How are they related?

[1] "Consideration" is the usual fudge-word for the kind of thing which can be a reason. See, for example, the use of Scanlon [1998]. But there are good reasons to think that properly speaking, considerations are something like facts or true propositions.

[2] This has been prominently, although not exactly explicitly, disputed. See Nagel's account of reasons in *The Possibility of Altruism*. Nagel appeals to claims of the form, "there is a reason for Ronnie to go to the party" in his account of claims of the form, "*R* is a reason for Ronnie to go to the party." So although he doesn't say so explicitly, he is obviously committed to holding that "there is" in "there is a reason for Ronnie to go to the party" is not an existential quantifier.

A little bit of vocabulary, then. Because they differ with respect to whether they have an agent as one of their relata, I'm going to say that the dyadic relation expressed by **1** is the *agent-neutral* reason relation, and that the triadic relation expressed by **3** is the *agent-relational* relation. So the reason referred to in **1**, because it stands in this relation, counts as *agent-neutral*, and similarly that in **3** counts as *agent-relational*. And again, I will use these same categories to refer to the *sentences* that *ascribe* agent-neutral and agent-relational reasons. So **1** itself is an agent-neutral *ascription*, while **3** is an agent-relational one. The question before us is: what is the relationship between agent-neutral and agent-relational reasons?

This is not, exactly, the same question as that of what the relationship is between agent-*relative* and agent-neutral reasons. According to the official definition, an agent-*relative* reason is a reason that is a reason for some people, but not for everyone, while an agent-*neutral* reason is one that is a reason for everyone. On this official definition, the distinction between agent-relative and agent-neutral reasons is one *among* agent-*relational* reasons. So according to the official usage, a reason cannot be both agent-relative and agent-neutral. Agent-relative reasons, then, are by definition *merely* agent-relational reasons.

2.2 The Quantification Strategy

The official definition takes a stand on the relationship between agent-relational and agent-neutral reasons. It claims that agent-relational reasons are basic, and that agent-neutral reasons arise when something is an agent-relational reason for everyone. I like this view, and call it the *Quantification Strategy*. According to the Quantification Strategy, the connection between **1** and **3** is a little bit like the connection between **1** and **2** or between **3** and **4**. Except instead of involving an *existential* quantifier, the connection involves a *universal* quantifier. The Quantification Strategy is basically that **1** involves a universal quantification into the agent-place of **5**:

5 The fact that Katie needs help is a reason for *x* to help Katie.

The Quantification Strategy is a very common view, and many philosophers think that it is obvious. I agree with these philosophers that something much like it is the *best* view to take about this question, and it is half of the purpose of this paper to explain why. But I do not agree that it is obvious. I think, in fact, that it is highly controversial. The fact that it is controversial is evidence that it is not obvious, and the fact that a number of interesting philosophical views involve rejecting it is evidence that it is controversial.[3] The other half of the purpose of this paper is to illustrate how interesting

[3] And the fact that it is controversial shows that we need a more theory-neutral way of distinguishing agent-relative and agent-neutral reasons than the official definition. I've just given such a theory-neutral way of drawing the distinction.

views about a variety of other questions can be supported by taking *other* views about the relationship between agent-relational and agent-neutral reasons.

Before I explain what the views *are* which conflict with the Quantification Strategy, however, let me explain why the Quantification Strategy *is* such a good theory about the relationship between agent-relational and agent-neutral reasons. Any adequate answer to this question is going to have to satisfy two central constraints:

Desideratum 1: Necessarily, for all *r* and *a*, if *r* is an agent-neutral reason to do *a*, it follows that for all agents *x, r* is an agent-relational reason for *x* to do *a*.
Desideratum 2: It is not the case that for all *r, x*, and *a*, if *r* is an agent-relational reason for *x* to do *a*, then *r* is an agent-neutral reason to do *a*.

If Katie needs help, then there is a reason to help Katie. And if there is a reason to help Katie, then it is a reason for me to help Katie and a reason for you to help Katie. Indeed, it is a reason for anyone to help Katie. That is what Desideratum 1 tells us. Of course, if the two of us like to dance, we *can* sometimes *say* things like "there is a reason to go to the party tonight" even though we know that there is not an agent-relational reason for Bradley to go there. But these cases are not counterexamples to Desideratum 1, because in such cases we don't mean to be asserting the existence of a genuinely agent-neutral reason. Though we don't mention an agent when we make such assertions, our assertions are merely elliptical. The relation they invoke is manifestly agent-relational, and we are merely saying that there is a reason *for us* to go to the party. All Desideratum 2 says, on the other hand, is that some agent-relational reasons, such as Ronnie's, are *merely* agent-relational.

It's easy to see why the Quantification Strategy is such a natural response to these two desiderata. It trivially satisfies a considerable strengthening of Desideratum 1. And it is clearly compatible with Desideratum 2. Of course, someone could accept the Quantification Strategy and then go on to give a bizarre account of the agent-relational reason relation from which it followed that if *R* is a reason for *X* to do *A*, then for all *x*, *R* is a reason for *x* to do *A*. But such an account would be treating the agent-relational reason relation as if it had no agent place at all. Any account of that relation whatsoever that made the agent-place non-trivial would satisfy Desideratum 2.

The Quantification Strategy, moreover, also has a number of other virtues. Most notably, it lets us treat all agent-neutral reason *ascriptions* on a par. I've just noted that we can sometimes say "there is a reason to go to the party" and only mean that there is a reason for *us* to go there. But if that is right, then we can think of the Quantification Strategy as simply a special case of this more general phenomenon. In general, we can say that:

Generalized QS: *R* is a reason to do $A \equiv_{def}$ *R* is an agent-relative reason for all of [us] to do *A*.

In some contexts—especially moral contexts—like those in which we can say things like "the fact that Katie needs help is a reason to help her," the context can determine that the scope of who counts as one of "us" must include everyone—even

every possible agent. And then when we talk in this way we will be talking about agent-neutral reasons. But in others, like that in which we are discussing our plans for this evening, only some agents—we ourselves—are salient in the context, and Bradley is not.

So the word "reason" and the function of the "there is a reason to" construction isn't simply ambiguous across those cases, even though in one we manage to ascribe the agent-neutral reason relation and in the other we do not. The required difference arises simply as a result of how we restrict the scope of who counts as one of *us*. And this seems like an attractive result, over and above the required desiderata which any theory has to satisfy. It suggests, I think, that we are on the right track.

2.3 Rationalism and the Inversion Strategy

Yet as natural as the Quantification Strategy seems, and as obvious as many find it, it is equally obvious that there are very interesting and important views in moral philosophy which are committed to rejecting it. Consider a version of the age-old question of moral philosophy, "why should I be moral?": "why is there a *reason* for me to be moral?" According to a traditional Rationalist kind of response to this question, the answer is, "because *there is* a reason to be moral." Christine Korsgaard tells us[4] that this answer amounts to mere foot-stomping: "Having discovered that obligation cannot exist unless there are actions which it is necessary to do, the realist concludes that there are such actions." But we can do better for the Rationalist than this. After all, the Rationalist is not so silly as to say that there is a reason for you to be moral because there is a reason *for you* to be moral. What he says is that there is a reason for you to be moral because *there is a reason to be moral*. So on the face of it, he is not stomping his foot at all, but suggesting that your agent-*relational* reason to be moral is explained by an agent-*neutral* reason to be moral.

This answer to the problem can seem hard to credit, if we accept the Quantification Strategy. For if we accept the Quantification Strategy, then we think that there being an agent-neutral reason to be moral simply *consists* in there being an agent-relational reason for each person to be moral. So we *do* think that the Rationalist answer sounds merely like foot-stomping. It sounds a bit like saying that you have a reason to be moral because you *and Alice both* have reasons to be moral. But if it sounds like that to us, that is only because we are convinced by the Quantification Strategy. The Rationalist, I'm suggesting, if we are to take his answer to the question, "why is there a reason for me to be moral?" seriously, has to be understood as holding that agent-neutral reasons

[4] Korsgaard [1996, 34]. Korsgaard goes on to say what she thinks is problematic about this kind of Rationalism, which she called "realism": "In a similar way, if someone falls into doubt about whether obligations exist, it doesn't help to say, 'ah, but indeed they do. They are *real* things' " [38]. And again: "And that is the problem with realism: it refuses to answer the normative question. It is a way of saying that it can't be done" [39].

explain agent-relational reasons, rather than conversely.[5] Since Rationalism of this kind is a possible view, gives us a very interesting answer to one of the central traditional questions of moral theory, and is committed to denying the Quantification Strategy, the Quantification Strategy can't be uncontroversial.

Rationalists of this stripe are committed to *inverting* the order of explanation of the Quantification Strategy. They hold that the agent-relational reason relation needs to be understood in terms of the agent-neutral one, rather than conversely. Let us call all views of this kind versions of the *Inversion Strategy*. The Rationalist needs a particular version of the Inversion Strategy, which I will explain in sections 2.8, 2.9, and 2.10. But first, let's look at two more versions of the Inversion Strategy. Each of these is useful to another important moral theory, just as the Standard Model is useful to this kind of Rationalism.

2.4 Kantians and the Point of View Theory

Rationalism of this kind is perhaps the view most obviously committed to rejecting the Quantification Strategy. But it is not the only such view. Consider an updated version of Kant's argument that any rational agent is required to abide by the Categorical Imperative. According to Kant, this commitment must be derived from the bare concept of rational agency, which involves only the idea of acting according to laws or, as an updated Kantian might put it, *reasons*. To be an agent, you have to see yourself as acting for reasons—considerations which are, in fact, reasons to act. But if you have to see yourself as acting in accordance with reasons in favor of one action or another, then we can apply Desideratum 1, which tells us that if there is a reason in favor of some action, it must be a reason for *each* person to perform that action. So it follows from the Kantian argument that in order to be an agent, you have to see yourself as acting for reasons *that are reasons for everyone, and not just you*—i.e., you have to be able to conceive your actions as in accordance with *universal* reasons, or, substituting back into Kant's language, universal *laws*. And that is just Universal Law formulation of the Categorical Imperative, or as close to it as I can find an argument.

But there is one particularly worrisome step in this argument. And I don't mean the step which claims that in order to act for reasons, you have to see yourself as acting for things which you take to be reasons, although I find that also questionable. I mean the assumption that when you take yourself to be acting for a reason, it must be an agent-neutral reason that you take yourself to be acting for. Why couldn't you

[5] If this version of the Rationalist's answer to the question "why should I be moral?" doesn't sound familiar to you, try the version in which the questioner asks, "why ought I to be moral?" and the Rationalist responds that being moral is *required* or *right*. Here, again, he appears to be using an agent-neutral normative concept in order to explain an agent-relational one. And again, many would differ. Many would think that what it is for an action to be right is just for it to be the case that everyone ought to do it.

take yourself to be acting for an agent-relational reason? One might think that if you are Ronnie, and you are clear-headed, you can go to the party for the reason that there will be dancing there, even though you are perfectly well aware that it isn't a reason for Bradley to go there. At most what you have to do, is to recognize that it is a reason *for you* to go there. Nothing more, it seems, could be required by the mere concept of rational agency.

But it's not hard, I think, to supplement the Kantian argument with a view about agent-relational reasons which rules this possibility out—a view which, however, conflicts with the Quantification Strategy. And the view is one which it is hard not to find tempting to attribute to a variety of Kantian thinkers.[6] It is what I call the *Point of View* theory. According to the Point of View theory, the "for Ronnie" in "there is a reason for Ronnie to go to the party" works like a propositional operator on the proposition that there is a reason to go to the party. So saying that there is a reason for Ronnie to go to the party is like saying that *from Ronnie's point of view* it is the case that there is a reason to go to the party. This conflicts with the Quantification Strategy because the Quantification Strategy accounted for the agent-neutral reason relation in terms of the agent-relational one. But the Point of View theory accounts for the agent-relational relation in terms of the agent-neutral one.

The Point of View theory is neutral about how to interpret what "points of view" are. But imagine, as suggested by the "point of view" terminology, that a necessary condition on its being the case from Ronnie's point of view that *p* is that Ronnie *believes* that *p*. This would do wonders for the Kantian argument. Suppose, then, that you believe that there is a reason for you to go to the party, but not an agent-neutral reason to go to the party. This is what would have to be the case in order to create trouble for the Kantian argument. Then you are committed, by this version of the Point of View theory, to the claim that you believe that there is an agent-neutral reason to go to the party, but at the same time you actually believe that there is no agent-neutral reason to go to the party. And that, though not a contradiction outright, is at least a *Moorean* contradiction. So if you are to avoid Moorean contradictions, then you can't consistently act for reasons unless you take the reasons for which you act to be agent-neutral reasons—to be reasons equally for everyone. And that would give us the Formula of Universal Law. This doesn't prove that any Kantian is committed to the Point of View theory—perhaps my reconstruction of the "updated" argument is nothing like the version any Kantian has ever accepted—but it does illustrate how *useful* it would be to a Kantian.

[6] Korsgaard, most obviously. See, in particular, Korsgaard [1997b]. It is also tempting to read Scanlon this way, when he insists that a reason is a consideration which "counts in favor" of some attitude. Scanlon must not accept the account outlined in section 2.3, since he thinks that reasons count in favor of attitudes. But he also must not accept the Quantification Strategy, because then he would have to say that a reason merely counts in favor of an action, *relative* to some agent. See Scanlon [1998].

2.5 Consequentialism and the Propositional Theory

Kantians and Rationalists are still not the only ones to find denying the Quantification Strategy useful. The simplest alternative to the Quantification Strategy derives from the idea that the reason relation is not a relation between a consideration and an *action*, but between a consideration and a *proposition*. On this view, to say that there is a reason for Ronnie to go to the party is to say that there is a reason in favor of the following *proposition*: that Ronnie goes to the party. According to some who take this view,[7] the "for Ronnie" in "there is a reason for Ronnie to go to the party" does double work. It both qualifies the reason, which is agent-relational, *and* qualifies the proposition the reason is in favor of. So it is like saying that there is a reason for Ronnie in favor of the proposition that Ronnie goes to the party.[8] But according to others, we don't need the "for Ronnie" to do double work. And this gives them a way of accounting for apparently agent-relational reasons in terms of agent-neutral ones. For there to be a reason for Ronnie to go to the party, on this view, is simply for there to be an agent-neutral reason in favor of the proposition that Ronnie goes to the party. Call this the *Propositional Theory*.

The Propositional Theory is particularly useful for *Consequentialists*. Consider one of the chief differences between Consequentialism and common-sense morality. According to common sense, it can sometimes, at least in principle, happen that you ought not to do something, even if by doing it you can prevent two or more other people from doing exactly the same thing. Such cases are called *agent-centered restrictions*. But according to Consequentialists this cannot happen. Indeed, according to some Consequentialists, it is quite hard to understand how there could be such things as agent-centered restrictions. The Propositional Theory facilitates one straightforward argument that there couldn't be.

So suppose that by doing some wrong action, you can prevent two or more other people from doing exactly that action. Since the action is wrong, there is a reason for you not to do it. And since by stipulation it is exactly the same action, there are reasons for each of the other two or more people not to do it. And these reasons are equally good. By the Propositional Theory, this means that there is an agent-neutral reason against your doing the action, and an agent-neutral reason against each of the others performing this action, and these reasons are equally good. But since the reasons are agent-neutral, they bind you equally. One weighs against your doing the action, but at least two weigh in favor of it. So the weight of reasons weighs in favor of your doing the wrong action. And so, concludes the argument, this must be what you ought to do.

This is an elegant argument against the possibility of agent-centered restrictions. But the Propositional Theory was crucial to running the argument. If we take the view that

[7] For example, according to John Broome [2004, 28–31].

[8] I explain some of the hard problems for the analogous view about "ought" in Chapter 7, this volume.

reasons weigh in favor of *actions*, rather than *propositions*, then agent-centered restrictions give us no trouble. In that case, we agree that if the action is *wrong*, then there is a reason for each agent not to do it. So there is a reason *for you* not to do it, and a reason for each of the people who you could prevent, not to do it. And the reasons can all be equally good, but the reasons for the others not to do it do not translate into equally good reasons for you to prevent them from doing it. For they aren't agent-neutral reasons in favor of the *state of affairs* that they not do it, but agent-*relational* reasons *for them* not to do it. It is only the Propositional Theory which makes us conflate these two things. So on this view it makes perfect sense to think that you still ought not to perform the wrong action.

There may be other and better reasons to be puzzled about agent-centered restrictions; I can't evaluate that here.[9] But it should be clear that the Propositional Theory would be *very useful* to the Consequentialist. For the Propositional Theory definitely validates this argument for the impossibility of agent-centered restrictions. And I find it hard not to think that this argument is something like what philosophers who are skeptical about agent-centered restrictions often have in mind, at least implicitly. So once again we can see that an issue important to a substantive and interesting moral theory is tied up with taking a view about the connection between agent-relational and agent-neutral reasons which conflicts with the Quantification Strategy.

We now know about three very interesting substantive issues in moral philosophy which are intimately related, to one degree or another, to views about the connection between agent-relational and agent-neutral reasons that conflict with the Quantification Strategy. Such views are all versions of the *Inversion Strategy*. So now let us turn to developing these versions of the Inversion Strategy enough to see whether they are, in fact, viable views. Though I have tried in sections 2.3, 2.4, and 2.5 to argue that they are *substantive* and *interesting*, I will now explain why they are, nevertheless, inadequate, and in ways that are quite independent of their implications discussed in sections 2.3, 2.4, and 2.5.

2.6 What's Wrong with the Propositional Theory?

So let's take the three versions of the Inversion Strategy in reverse order, from easiest to hardest, and the Propositional Theory first. The Propositional Theory, I think,

[9] Well, okay, but just one pass. There are certainly other *reasons*, but I'm skeptical about whether they are better. For example, the argument might turn on the assumption of some kind of *teleological* explanation of the reasons involved. Such arguments often do work this way—the proponent tries to convince us that if the action is wrong for you and each of the others to do, then that must be because there is something *bad* about any of you doing that action. But surely this kind of teleological assumption should be more obviously controversial in this context. Deontology is often characterized as precisely the view which rejects this kind of teleological explanation of reasons in terms of value or disvalue. So since the teleological argument has more obviously controversial premises, I take it that it is not a better motivation for the Consequentialist's view. See Scheffler [1982], Kagan [1989], and Smith [2003].

is clearly the worst of the available views. It relies, for starters, on the false view that reasons are in favor of *propositions* instead of *actions*. And it still fails by the lights of our two desiderata on any answer to our original question, because it has no plausible candidate for what the proposition is that the reason is in favor of in our original paradigmatic agent-neutral reason ascription, "there is a reason to help Katie."

Set aside the question of whether reasons weigh in favor of actions or propositions.[10] Let's just look at how the Propositional Theory is to account for our two desiderata. It's easy to see what the proposition is that the reason is in favor of, in the case of "there is a reason for Ronnie to go to the party." It is the proposition that Ronnie goes to the party. But what is the proposition that the reason is in favor of in the agent-*neutral* ascription, "there is a reason to help Katie"?

We want Desideratum 1 to turn out to be true, so consider the candidate that it is the proposition that *everyone* helps Katie. If there is a reason in favor of the proposition that everyone helps Katie, then it is certainly a reason in favor of the proposition that you help Katie. After all, your helping Katie is necessary for it to be the case that everyone helps Katie. So that would explain Desideratum 1. But clearly there can be a reason to help Katie even if there is no reason in favor of the proposition that *everyone* helps Katie. After all, if Katie needs help, she doesn't need *everyone* to help her—she just needs *enough* help. So this view about what proposition the agent-neutral reason is in favor of requires something far too strong.

On the other hand, perhaps the reason to help Katie is a reason in favor of the proposition that *someone* helps Katie. But then why should we think that there is a reason for *each* person to help Katie? In particular, why should we think that there is a reason in favor of the proposition that *you* help Katie? Why must a reason in favor of the proposition that *someone* helps Katie necessarily be a reason in favor of the proposition that *you* help Katie? This question is harder to answer, if we keep in mind just how plausible it is that a reason must count in favor of each of the necessary consequences of the proposition that it is in favor of. If that is right, then if there is a reason in favor of the proposition that Ronnie goes to the party, it must also be a reason in favor of the proposition that someone goes to the party. But then our explanation of your reason to help Katie only works if it *also* follows that Ronnie's reason is also a reason in favor of the proposition that *Bradley* goes to the party. And then the Propositional Theory fails on account of Desideratum 2.

The only way to save the Propositional Theory is to abandon the view that a reason must count in favor of the necessary consequences of the proposition it is in favor of. But the idea that the force of reasons transfers from ends to *necessary* means is one of the bedrocks of thinking about how the force of reasons transfers from one thing to another. So the Propositional Theory looks hopeless for accounting for the two most obvious central desiderata for any theory about the connection between agent-relational

[10] I've addressed the question for the case of "ought" in Chapter 7, this volume, and hope to take up further arguments for the view on a future occasion (Schroeder [2011]).

and agent-neutral reasons. Correlatively, Consequentialists should look harder for a better argument that there is something puzzling about agent-centered restrictions, and non-Consequentialists should check Consequentialist reasoning carefully, to see whether it covertly relies on the Propositional Theory—or anything very much like it.

2.7 What's Wrong with the Point of View Theory?

According to the Point of View theory, saying that there is a reason for Ronnie to go to the party is like saying that from Ronnie's *point of view*, there is a reason to go to the party. In principle, the Point of View theory isn't committed to thinking that *points of view* are anything like points of view in any colloquial sense, although the version of the Point of View theory required in order to make my updated Kantian argument work *did* require this. But in principle, all that the Point of View theory is in general committed to holding, is that the "*for Ronnie*" works like a propositional operator on a content that is about agent-*neutral* reasons.

Yet it is relatively easy to rule out a wide range of versions of the Point of View theory, on the grounds of our two desiderata for any viable answer to our original question. For according to Desideratum 1, it must follow from the fact that there is an agent-neutral reason to help Katie that there is an agent-relational reason for you to help Katie. So according to the Point of View theory this is like saying that if there is a reason to help Katie, it must be the case that *from your point of view* there is a reason to help Katie. And that is like saying that agent-neutral reasons are transparent to your point of view. And if that is so, then points of view can't be anything like points of view, in the conventional sense. Reasons certainly aren't transparent to our beliefs—otherwise how could there be so much disagreement about what the reasons are?

Compare again: in order to get the updated Kantian argument to work, I claimed that it had to be a *necessary* condition on its being the case that *p* from Ronnie's point of view that Ronnie *believed* that *p*. But Desideratum 2 requires that it be possible that there is a reason for Ronnie to go to the party, even though there is no reason to go to the party. So according to the version of the Point of View theory required in order to make the updated Kantian argument work, it follows that it must be possible that Ronnie sometimes believes that there is a reason to do something which there is no reason to do.

So put together, we have two constraints on how points of view must work, derived from our two desiderata. First, agent-neutral reasons must somehow be transparent to points of view. Yet there must also be *further* reasons from points of view which are not, in fact, actual agent-neutral reasons. These two constraints rule out any ordinary understanding of what points of view are that I know about, including that necessary in order to run the updated Kantian argument. I don't claim that this refutes the Point of View theory in general, but I do claim that it undermines much of the source of its intuitive appeal—the appeal of thinking that when you say that there is a reason for

Ronnie to go to the party, you are saying something like that for Ronnie, it is as if there is a reason to go to the party.

The challenge which remains for the still-standing versions of the Point of View theory is not simply to demonstrate *consistency* with our two desiderata on any viable theory. It is to provide an *explanation* of *why* these two desiderata, and particularly the first, are true. The Quantification Strategy isn't merely consistent with the two desiderata. It gave us a good explanation of *why* Desideratum 1 is true. For Desideratum 1 is a consequence of the Quantification Strategy. I've now argued that *some* plausible-sounding versions of the Point of View theory are inconsistent with the two Desiderata, and left open whether other versions are consistent with it. Yet a *good* theory about the relationship between agent-relational and agent-neutral reasons should actually *explain* the most central and obvious entailments connected with these two relations. *If* there is an adequate version of the Point of View theory—a possibility that I don't rule out—it will have to do this for us.

One final unpromising feature of the Point of View theory. If "for Ronnie" is simply a propositional operator, then it should in principle be able to be applied to contents with other sorts of content, not about agent-neutral reasons. For example, it should make sense to say that "there is a tunnel for Ronnie under the English Channel" or that "for Ronnie the Potomac floods in the spring." Whatever account the Point of View theorist gives us about this operator, it has to be one which makes these claims make as much sense at the one that "there is a reason for Ronnie to go to the party," in *addition* to explaining the two desiderata. I'm not optimistic.

2.8 The Subsumption Account

The third version of the Inversion Strategy—the one that is necessary for the Rationalist to run his response to the question, "why is there a reason for me to be moral?"—is more sophisticated than the two I've just set aside in sections 2.6 and 2.7, and I think many more people accept it, at least implicitly. It will also require much more care, in order to see what is deeply problematic about it. A first attempt at such an account might go like this: take the agent-neutral reason relation as relatively primitive, and say:

Subsumption #1: *R* is an agent-relational reason for *X* to do *A* $\equiv_{def} \exists b$ *R* is an agent-neutral reason to *b* & *A*-ing is a way for *X* to do *b*.

The thought behind this attempt is that if we are to account for the agent-relational reason relation in terms of the agent-neutral reason relation, then for each agent-relational reason we need some agent-neutral reason. In order to satisfy Desideratum 1, it had better be the *same* consideration, so we build that in to our account. But in order to satisfy Desideratum 2, we need some sort of leeway between what *R* is an agent-neutral reason to do, and what it is an agent-relational reason for someone in particular to do. *Ways* of doing something are supposed to do this work for us. In this way, we *subsume*

agent-relational reasons under agent-neutral reasons. There is an agent-relational reason for you to do something, when it is a *way* of doing something that there is an agent-neutral reason to do.

So, for example, there may be an agent-neutral reason to eat healthily. But what it takes to eat healthily may differ from agent to agent. Diabetics and Atkins dieters need to eat in different ways from the rest of us, in order to eat healthily. So it follows from Subsumption #1 there is an agent-relational reason for Diabetics to eat in ways that there is no agent-relational reason for Atkins dieters to eat, and vice versa. We can think of *ways* talk as a generalization on *means–end* talk, if that's helpful.

This account actually does remarkably well. So long as we assume (as we should) that *A*-ing is always a way for *X* to do *A*, the account satisfies Desideratum 1. Moreover, so long as we assume that it is sometimes the case that *A*-ing is a way for *X* to do *B*, but not a way for *Y* to *B*, the view also satisfies Desideratum 2. Yet it still gets something deeply wrong. To see why, we need to look more carefully at how this account will deal with the case of Ronnie and Bradley (Bradley, recall, is the one who doesn't share Ronnie's merely agent-relational reason to go to the party, because he can't stand dancing.) It must find some action-type *B* such that going to the party is a way for Ronnie to do *B*, but not for Bradley. Let's suppose for the sake of argument that this action is *doing what one likes to do*. The account will assume that there is some $\mathfrak{R}$ such that $\mathfrak{R}$ is an agent-neutral reason to do what one likes to do. Let pass for now exactly what $\mathfrak{R}$ might be. Whatever it is, according to this account, $\mathfrak{R}$ is the agent-relational reason for Ronnie to go to the party. Intuitively, however, the reason for Ronnie to go to the party is that there will be dancing there—something that is special to Ronnie's case. Intuitively, Ronnie has some reason to go to the party that is not the same as Brett's reason to practice his bass or Vera's reason to play chess, things that Brett and Vera like to do.

The problem is that though this account vindicates claims of the form, "there is a reason for *X* to do *A*," it fails miserably to capture truths of the form, "*R* is a reason for *X* to do *A*." Because "there is a reason for *X* to do *A*" transparently quantifies[11] over reasons, we really are committed to things being reasons. So our theories had better be able to account for those reasons. This point is usually ignored in the literature, but that doesn't make it irrelevant.[12] So this first attempt at the inversion strategy yields unacceptable results.

A second attempt turns out to be less satisfactory, but at least to start us in the right direction. The idea is this: the first account got us into trouble because it made all of the agent-relational reasons that were accounted for using the same agent-neutral reason

[11] On the face of it, Thomas Nagel denies this (see section 2.1). But it's not as though he denies that there are things that are reasons—on the contrary, he commits to an account of what sort of things they are. Much argument would be required to establish that we shouldn't think that when there is a reason to do something, something is the reason to do it.

[12] Ignoring what someone's reasons *are* to do something *can* be a way of simplifying discussion of some related philosophical questions. But some of the issues that doing so glosses are important, and this is one. We can't simply ignore it all of the time.

come out to be the same—and the same as that agent-neutral reason. So we should instead build into our account a way for different agent-relational reasons to derive from the same agent-neutral reason, in a way that will reflect our intuitive judgments about what Ronnie's, Brett's, and Vera's reasons are. The account that gets this part roughly correct says:

Subsumption #2: *R* is an agent-relational reason for *X* to do *A* $\equiv_{def}$ $\exists s\ \exists b$ *s* is an agent-neutral reason to do *b* & *A*-ing is a way for *X* to do *b* & *R* is an essential part of the explanation why *A*-ing is a way for *X* to do *b*.

Why is going to the party a way for Ronnie to do what he likes? Well, it's because (1) he likes to dance, and (2) there will be dancing at the party. Each of these facts figures essentially in this explanation, and each is intuitively a reason for him to go to the party. So this account explains what is going on in Ronnie and Bradley's case, as long as we again assume that there is some $\mathfrak{R}$ which is the agent-neutral reason to do what one likes.

Unfortunately, however, this account fails by the lights of Desideratum 1. It would seem that nothing figures essentially in the explanation of why helping Katie is a way for *X* to help Katie—after all, this is trivial. But then it follows that nothing *at all* about agent-relational reasons follows from the fact that there is an agent-neutral reason to help Katie—namely, that she needs help. Nothing turns out to be a reason for anyone to help Katie, in virtue of the existence of some agent-neutral reason to help Katie. So this account not only gets wrong what various agents' agent-relational reasons are, it also yields the wrong results about when they have agent-relational reasons. In this regard, it is worse than the previous account. It does, however, help to illustrate what motivates the account in the next section, and to show why the Subsumption Account needs to resort to such a view.

2.9 The Best Subsumption Account

The trouble with the Subsumption Account is that if it is to validate the Rationalist's view, then for every agent-relational reason there must be some agent-neutral reason from which it derives in some way. Desideratum 1 puts pressure on the account to allow that the agent-neutral reason should just be the same consideration as the agent-relational reason that derives from it. But this guarantees the wrong results when we look more closely at Ronnie and Bradley's case. What we need is a way of allowing that these considerations should be the same in just those cases to which Desideratum 1 applies, but different in all of the cases which will worry us with respect to Desideratum 2. There is a unified way of getting this result, which does considerably better than the first account we considered, and it borrows from each of the previous two accounts. It says:

Subsumption #3: *R* is an agent-relational reason for *X* to do *A* $\equiv_{def}$ $\exists s\ \exists b$ *s* is an agent-neutral reason to do *b* & *A*-ing is a way for *X* to do *b* & *R* is the conjunction of *s* with each truth that figures essentially in explaining why *A*-ing is a way for *X* to do *b*.

This account simply takes as *X*'s agent-relational reason to do *A* the conjunction[13] of all of the things that counted as *X*'s agent-relational reasons to do *A* according to each of the last two accounts.

If we continue to suppose that nothing figures essentially in explaining why helping Katie is a way for *X* to help Katie, then this account satisfies Desideratum 1. The fact that Katie needs help is an agent-neutral reason to help her, and we don't have to conjoin anything with it in order to get an agent-relational reason for *X* to help her. Meanwhile, if we continue to suppose that there is some ℜ that is an agent-neutral reason to do what one likes, then the account can also do tolerably well with respect to Ronnie and Bradley's case. ℜ conjoined with the fact that Ronnie likes dancing and with the fact that there will be dancing at the party is an agent-relative reason for Ronnie to go there.

Of course, what we wanted was that each of the latter was an agent-relational reason for Ronnie to go there, and not merely that their conjunction with the unknown ℜ would be. So even on this account, our ordinary ascriptions of reasons are not literally true. It improves on the first inversion account by allowing that Ronnie and Brett and Vera all have different agent-relational reasons, but it doesn't quite make the fact that there will be dancing at the party turn out to literally be a reason for Ronnie to go there. All it yields is that this is one of the conjuncts in a reason for Ronnie to go there, or part of one of Ronnie's reasons to go there.

But this may be good enough. True, sometimes we can say that the fact that there will be dancing at the party is a reason for Ronnie to go to the party, and sometimes we can say that the fact that Ronnie likes dancing is a reason for him to go to the party. But when Ronnie is deciding to go to the party, it seems, he shouldn't add the weights of these two reasons together—they seem to count in the same way towards his going to the party. So perhaps there really is some sense in which they are really the same reason.[14] The third version of the inversion strategy gets this intuition right. According to it, each ascription functions to ascribe literally the same reason, surface appearance to the contrary.

There's a familiar idea from the literature which explains why we can use ascriptions like "the fact that there will be dancing at the party is a reason for Ronnie to go to the party" and "the fact that Ronnie likes dancing is a reason for Ronnie to go to the party" in order to ascribe the same reason.[15] It is that contextual conversational

[13] Another version of this same account might not appeal to *conjunction*. It might make the reason out to be the fusion of each of the other facts, or something like that. It's all the same, for what I have to say.

[14] I don't myself hold this view; I'm merely articulating why things don't look so bad for Subsumption #3. I've articulated my contrasting view in Schroeder [2009a] and in Schroeder [2007a].

[15] See, for example, Davidson [1980]. Davidson isn't talking in that paper about normative reasons—the kind of reasons in which we're interested—but rather about *motivating* reasons. Still, he didn't distinguish clearly, and neither did many who followed him. As a result, it's not uncommon to see philosophers supposing without argument that someone's normative reason is really the entire sufficient condition for it to be the case that she has a reason. This is what Thomas Nagel says, for example, in *The Possibility of Altruism* [1970]. But it's worth noting that pragmatic contextualism about reason ascriptions is also compatible with, though not a commitment of, the Quantification Strategy. The Quantification Strategy, however, *can* make the different reason ascriptions all come out to be literally true.

pragmatics[16] determines which information about someone's reason is needed in the conversation. If we already know that Ronnie likes dancing, then it makes sense to say that the fact that there will be dancing at the party is a reason for him to go there. Even though this isn't literally true, we have everything that we need in order to construct what Ronnie's complete reason really is, so the rest isn't worth mentioning. It may also be true that we can *infer* that this fact would only be citable as a reason for him to go there, if he liked dancing. In that case, though we do not know it already, we can infer that Ronnie likes dancing, and so this part of his reason again turns out not to be worth mentioning. Likewise, when only the fact that Ronnie likes dancing is mentioned, we either already know, or are able to infer, that there must be dancing at the party. That is why we don't have to mention them in ordinary-language reason ascriptions.

So as far as I can see, the Subsumption Account is going to have to be committed to some such contextualist pragmatics, in order to make sense of our actual reason ascriptions. This isn't a bad thing. As I noted, it helps to make sense of the idea that the fact that there will be dancing at the party and the fact that Ronnie likes dancing shouldn't count separately, for him, in favor of going there. But I've now argued that the Subsumption Account is committed to more than the claim that these are parts of Ronnie's reason. It is committed to the claim that the mysterious $\mathfrak{R}$, which is the agent-neutral reason to do what one likes, is also part of Ronnie's reason. Not only that, but in order for its conversational pragmatics to work, it is committed to holding that it is so obvious what $\mathfrak{R}$ is, that it is not worth mentioning. And it is this mysterious agent-neutral reason that turns out to make the most trouble for inversion views, and for this last and most promising inversion view, in particular.

2.10 A Mysterious Reason

So far, we've been assuming that inversion strategies can appeal to the existence of some $\mathfrak{R}$ that is an agent-neutral reason to do what one likes, in order to account for merely agent-relative reasons that arise from likes and dislikes. Unfortunately, making good on this assumption is not easy. It's hard to see what single feature of the world it could be that counts, for each person, in favor of that person's doing what she likes.[17]

[16] Indeed, the idea is so familiar that some philosophers have corrected me, when I have said things like, "intuitively, the fact that there will be dancing at the party is a reason for Ronnie to go there," and told me that this is obviously only part of Ronnie's reason. These philosophers find the contextualist pragmatics so obvious that they've ceased to see the benefit of an account that makes ordinary ascriptions literally true.

[17] Actually, it's also hard to characterize what the action could be that it is a reason to do, such that every action that needs to be explained by this reason really gets subsumed by it, *and* leads to a satisfactory explanation of the resulting reason. See Chapter 3, this volume, for details. Here, I'm only concerned with what $\mathfrak{R}$ is supposed to be, because in order to get the conversational pragmatics to work, the Subsumption View has to assume not only that there *is* such an $\mathfrak{R}$, but that we all *know* what it is—that it is so obvious as not to be worth mentioning.

It's not enough to assure ourselves that it's obvious that "everyone has a reason to do what she likes." For from

Transparent: $\forall x \forall a$ If x likes to do a then $\exists r$ (r is a reason for x to do a).

it certainly doesn't follow that

Opaque: $\forall x \exists r$ r is a reason for x to do what x likes to do.

let alone that

Needed: $\exists r \forall x$ r is a reason for x to what x likes to do.

Transparent says that when each person likes to do something, there is a reason for her to do *it*. Opaque says that each person has a general reason to do the following action (if there is such a thing): do what she likes. Needed says the same as Opaque, but asserts that this is an agent-neutral reason, the same for everyone. Transparent gives us an obvious reading of "everyone has a reason to do what she likes" that is obviously true.

But the truth of Transparent is compatible with there being no *general* agent-relational reason, even just for Ronnie, to do whatever he likes. It certainly does not entail the truth of the Opaque reading of the intuitive, "everyone has a reason to do what she likes." Moreover, even if it did, Opaque is compatible with each person having different agent-relational reasons, so that there is no single consideration to *be the* agent-neutral reason to do what one likes. Yet this is what the Subsumption Account requires. So let me illustrate:

The truth of the matter is that the fact that there will be dancing at the party might be an agent-relational reason for Ronnie to go there—and going there, after all, is doing what he likes. The fact that practicing his bass will improve his chances of becoming a rock star is a reason for Brett to practice his bass—and improving his chances of becoming a rock star, after all, is what he likes. And so on. The fact that Plato is a philosopher is a reason for Brett to study Plato—after all, Brett likes to study philosophers. Brett's reason to study Plato and his reason to practice his bass are both reasons that he has in virtue of what he likes, but they aren't the same reason. So it would follow that everyone has an agent-relational reason to do each of those things which happen to be among the things that she likes to do, but not that there is any general agent neutral reason to do what one likes. This makes sense of the intuitively obvious claim without appeal to an agent-neutral reason such as $\mathfrak{R}$. The claim *is* perfectly true—*on* the *transparent* reading, the one on which the description, "what she likes," takes *wide* scope over the reason-ascription.

Now there *may* be a *sense* in which, when Ronnie has a reason to go to the party because he likes to dance, and Brett has a reason to practice his bass because he likes getting better at playing his bass, Ronnie and Brett have the same reason. If there is, it is the same weak sort of sense in which, when Ronnie believes that he is in New York, and Brett believes that *he* is in New York, they believe the same thing. There is even a sense in which Ronnie and Brett have their reasons *because* they are the "same" in this way.

But this isn't enough for the Subsumption Account to work. Even the Quantification Strategy can allow that Ronnie and Brett have these reasons because there is some true generalization of the form,

True Generalization: $\forall x \forall a \forall b$ If x likes to do b and a-ing is a way for x to do b, the fact that a-ing is a way for x to do b is an agent-relative reason for x to do a.

And even the Quantification Strategy can allow that this is *why* the fact that Ronnie likes to dance is a reason for him to go to the party. For this might be part of the account of what it takes to have an *agent-relational* reason to do something. But that doesn't mean that the Quantification Strategy is appealing to a further, *agent-neutral*, reason in explaining Ronnie and Brett's agent-relational reasons. It's just allowing some true generalizations about what agent-relational reasons they have.

No matter what $\mathfrak{R}$ turns out to be, it is going to have to be a truth, a fact about the world, in the way in which the fact that Katie needs help is a fact about the world, and the fact that there will be dancing at the party is a fact about the world. And it has to be the sort of thing that intuitively counts, for each and every person, in favor of her doing what she likes. I simply have no idea what this fact is supposed to be. That doesn't mean that there's no such thing; it just means that the Subsumption Account owes us something significant, in order to make good on its promise to deal adequately with the case of Ronnie and Bradley.[18]

The going version of the Subsumption Account, however, is committed to more than this. For it is committed to some kind of contextualist pragmatics, in order to explain why it's not necessary for us to mention *all* of Ronnie's reason, in order to properly ascribe something as a reason for him to go to the party. Now all of the stories like this that I know of work by explaining why listeners in the context are able to infer the rest of the information about what Ronnie's reason is, without needing to be told. And this works quite well, so long as we only suppose that the fact that Ronnie likes to dance and the fact that there will be dancing at the party make it into Ronnie's agent-relational reason to go to the party. But in fact, in order to solve the problems with the second inversion account, we had to make sure that $\mathfrak{R}$, the agent-neutral reason to do what one likes, *also* makes it into his agent-relational reason, and in the same way as these

[18] Some possible candidates which I find unpromising: $\mathfrak{R}$ is the fact that sometimes doing what one likes is part of the *good life* for human beings. Or perhaps $\mathfrak{R}$ is the fact that other things being equal it is better that more people do what they like more of the time. If we say the former, we can't say that what makes the good life good is that there are reasons to prefer it, because then we're back in the water when it comes to explaining why this is the good life, looking for another agent-neutral reason to prefer a life in which one sometimes does what one likes. Likewise, if we say the latter, we can't say that what makes one state of affairs better than another is that there are sufficient reasons to prefer it, because then we're back in the water when it comes to explaining why such a state of affairs really would be better. Moreover, if we take the second line, we now have a puzzle about why Ronnie's reason to go to the party is any better than his reason to make sure that Bradley stays away. He can bring about just as much of people doing what they like by getting Bradley to stay away, after all, and on this account, his reason to do what he likes is just that it's better when more people do what they like.

other two facts. But according to the contextualist pragmatics, that should mean not only that there really is some $\mathfrak{R}$ that is a reason to do what one likes, but that it should be so obvious that we never need to mention it. And that seems to be a fairly strong commitment. It's certainly not obvious to *me* what $\mathfrak{R}$ is.

Still, suppose that it's obvious to everyone that there is *some* $\mathfrak{R}$ such that $\mathfrak{R}$ is an agent-neutral reason to do what one likes. Would that be enough to get the pragmatic contextualist account to work? Perhaps. The explanation couldn't be that the reason we don't have to mention $\mathfrak{R}$, however, is that we already know what it is. But that leaves room for other kinds of explanation. Still, I think the considerations in this section show that we need to be careful about whether, when it seems obvious that everyone has a reason to do what she likes, this is really what is being claimed. I truly do not find it obvious that there is such an $\mathfrak{R}$, and neither, I take it, should anyone else who accepts the Quantification Strategy. The real problem with the Subsumption Account, therefore, is that the contextualist pragmatics it needs seems to require that the Subsumption Account itself be uncontroversial. But nothing about the Quantification Strategy requires that the Quantification Strategy be uncontroversial. It just requires that it be true. And that is an undeniable theoretical advantage.

2.11 Concluding Remarks

In this paper, I've characterized the distinction between agent-neutral and merely agent-relational reasons in a theory-neutral way which, unlike the official definition of agent-neutral reasons, does not presuppose the controversial Quantification Strategy. I've defended a generalized version of the Quantification Strategy, and shown why it suits the criteria for an account of the relationship between agent-relational and agent-neutral reasons so well. And I've developed, and dismissed, three competing accounts—versions of the Inversion Strategy. But I've argued that the question is one of substance, not one to which to take answers for granted. It is all too often the taking for granted of answers to questions like this one, which underwrites deep disagreements about a variety of substantive issues in moral theory. And when that turns out to be the case, there is nothing for it but to investigate the adequacy of these views not by the lights of their controversial implications for ethical theory, but in terms of the most obvious constraints on the question to which they are supposed to provide answers. Sometimes, and I think in this case this is true, the *interestingness* of these views is not enough to overcome their theoretical inadequacies.[19]

[19] Special thanks to Gideon Rosen, Karen Bennett, Pete Graham, Joshua Knobe, Mark Johnston, Philip Pettit, Mike McGlone, Brett Sherman, Jeff Speaks, Ant Eagle, Nadeem Hussain, a very careful anonymous referee for *Philosophical Studies*, and to audiences of primitive versions of parts of this paper at both Princeton and NYU.

3

The Humean Theory of Reasons

This paper offers a simple and novel motivation for the Humean Theory of Reasons. According to the Humean Theory of Reasons, all reasons must be explained by some psychological state of the agent for whom they are reasons, such as a desire. This view is commonly thought[1] to be motivated by a substantive theory about the power of reasons to motivate known as *reason internalism*, and a substantive theory about the possibility of being motivated without a desire known as the *Humean Theory of Motivation*. Such a motivation would place substantial constraints on what form the Humean Theory of Reasons might take, and incur substantial commitments in metaethics and moral psychology. The argument offered here, on the other hand, is based entirely on relatively uncontroversial methodological considerations of perfectly broad applicability, and on the commonplace observation that while some reasons are reasons for anyone, others are reasons for only some. The argument is a highly defeasible one, but is supposed to give us a direct insight into what is *philosophically deep* about the puzzles raised for ethical theory by the Humean Theory of Reasons. I claim that it should renew our interest in the relationship between these two kinds of reason, and in particular in the explanation of reasons which seem to depend on desires or other psychological states.

3.1 The Humean Theory of Reasons: What

Consider a case like that of Ronnie and Bradley. Ronnie likes to dance, but Bradley can't stand even being around dancing. So the fact that there will be dancing at the party tonight is a reason for Ronnie to go there, but not for Bradley to go there—it is a reason for Bradley to stay away. Ronnie and Bradley's reasons therefore differ—something is a reason for one to do something, but not for the other to do it. And this difference between their reasons seems obviously to have something to do with their psychologies. It may not be ultimately explained by the difference in what they *like*, of course—the explanation may ultimately derive from a difference in what they *value*, or what they *care* about, what they *desire, desire to desire*, what they take or would take

[1] See, for example, Williams [1981], Bond [1983], Darwall [1983], Korsgaard [1986], Hooker [1987], Hubin [1999], and others.

pleasure in, or what they *believe to be of value*. I'm not claiming that it is uncontroversial that one rather than another of these kinds of psychological states is what really explains the difference between Ronnie and Bradley—after all, many of these psychological characteristics often go hand-in-hand, and even moderately sophisticated views can make them hard to distinguish simply by considering cases. All I'm claiming is that it should be pretty close to uncontroversial that there are *at least some* reasons like Ronnie's, in that they are explained by *some* psychological feature.[2]

The *Broad* Humean Theory of Reasons says that all reasons are explained in the same way as Ronnie's—by the same kind of psychological feature:

Broad Humean Theory: Every reason is explained[3] by the kind of psychological feature that explains Ronnie's reason in the same way as Ronnie's is.

The Broad Humean Theory of Reasons is really too broad to sound familiar to most readers familiar with the philosophical literature on reasons. That literature is full of references to, and attacks on, a familiar view that is more *narrow* than the Broad Humean Theory. This view is a *version* of the Broad Humean Theory because it agrees that all reasons must be explained by the same kind of psychological feature as explains Ronnie's. But it is more *specific* than the Broad Theory, because it takes a view about *what kind* of psychological state *does* explain the difference between Ronnie's and Bradley's reasons. It says that it is a *desire*, in the traditional philosophical sense:

Narrow Humean Theory: Every reason is explained by a desire in the same way as Ronnie's is.

Even the Narrow Humean Theory of Reasons, of course, is only loosely called "Humean"; there is an excellent case to be made that Hume himself was not a Humean in either sense. Both theories are associated with Hume's name primarily because their proponents have typically been loosely inspired by Hume.[4]

[2] Allow me to head off a possible distraction. There is a sense in which what reasons one has depends on what one believes. In this sense, though there will be dancing at the party and Ronnie and Freddie both like to dance, if Freddie is aware of this but Ronnie is not, then we might say that Freddie has this reason but Ronnie does not. This is the subjective sense of 'reason'. When I say that it is uncontroversial that at least some reasons depend on psychological states, this is not what I intend. What I mean, is that it is uncontroversial that at least some reasons in the objective sense depend on psychological states.

[3] A qualifying note about how to understand this talk about *explanation*. The fact that there will be dancing at the party tonight is a reason for Ronnie to go there, in part *because* Ronnie likes to dance. That must be part of *why* it is a reason for Ronnie to go there, because it is not a reason for Bradley to go there, and liking to dance is precisely what distinguishes Ronnie from Bradley. The Humean Theory of Reasons is a generalization of *this* claim. It is the claim that whenever *R* is a reason for *X* to do *A*, that is in part *because* of something about *X*'s psychology—that this is part of *why R* is a reason for *X* to do *A*. I'm using the term "explained by" to cover these kinds of claims about what is so *because* something else is so, and what is part of *why* it is so. This is not intended to import epistemic or pragmatic ideas about what *agents* might be doing when they engage in the behavior of *explaining* things to one another. In my sense, *X* explains *Y* iff *Y* is the case *because X* is the case, or *X* is part of *why Y* is the case. The explanation is the *content* of the answer to a "why?" question—not the answer itself, nor the process of giving it.

[4] So it's not worth quoting Hume for the purpose of refuting either view. Compare Korsgaard [1997b]. See also Setiya [2004] for an excellent discussion of how to understand Hume's commitments about practical reason.

So allow me to reveal my hand. I believe that a version of the Narrow Humean Theory of Reasons is true, and I have defended such a theory elsewhere.[5] But in this paper I will not be arguing for the Narrow Humean Theory. The argument of this paper is only a motivation for the Broad Humean Theory. It is my *view* that there are good arguments from the Broad Humean Theory to the Narrow Humean Theory, but I will not advance those arguments in this paper. Indeed, I think that for most of the philosophical reasons for which philosophers have been interested in whether the Humean Theory of Reasons is true, whether the Humean Theory is *Narrow* or not is beside the point. In the next subsection I will explain why.

3.2 The Humean Theory of Reasons and Moral Skepticism

The Broad Humean Theory of Reasons takes no stand on what kind of psychological state it is that does explain the difference between Ronnie and Bradley. It only claims that whatever it is, it is *also* needed to explain every other reason. But this does not water the Humean Theory down so much as to make it of little interest. On the contrary, it is exactly the right specificity of view that we should be worried about, for exactly the reasons that philosophers have been worried about the Narrow Humean Theory of Reasons all along.

The principal philosophical interest of the Narrow Humean Theory of Reasons, after all, is that it is supposed to play a special role in motivating certain kinds of skepticism about the universality or objectivity of morality. The problem is that according to the Humean Theory, every reason must be explained by a desire of the person for whom it is a reason. But it is hard to see how such an explanation could possibly work for all moral reasons. Consider this case: Katie needs help. So there is a reason to help Katie. It is a reason for you to help Katie, a reason for me to help Katie, and in general, it is a reason for *anyone* to help Katie. Some of the most important moral reasons seem to be like the reason to help Katie—they are reasons for *anyone*, no matter what she is like. But does *everyone* really have some desire that would explain a reason for her to help Katie in the *same* way that Ronnie's desire to dance explains his reason to go to the party? It seems fairly implausible.

So those who accept versions of the Narrow Humean Theory often take revisionist views about the kind of objectivity that moral claims have. Gilbert Harman, for example, argues for these reasons that moral claims aren't really universally binding, but are only binding on people who have implicitly contracted in certain ways. This is his brand of moral relativism in 'Moral Relativism Defended' and subsequently.[6] Philippa Foot argues for almost identical reasons that moral claims don't provide reasons to everyone, but only to those who care about morality. That is her thesis in 'Morality as a System of

[5] Chapter 7, this volume, Schroeder [2007a], [2007c].
[6] Harman [1975]. See also Harman [1978] and [1985].

Hypothetical Imperatives'.[7] The difference between Harman and Foot is that Foot thinks that there is another, non-reason-giving, sense in which moral claims nevertheless "apply" to everyone, even to those to whom they don't give reasons. John Mackie argues that it is essential to moral claims that moral requirements give reasons to everyone. Since this is incompatible with the Humean Theory of Reasons, he concludes that moral claims are uniformly false.[8] These are all drastic forms of skepticism about the objectivity or universality of morality that are motivated by the Humean Theory of Reasons. And it is these kinds of arguments which give the Humean Theory so much of its interest for moral theorists. It is in order to *avoid* these kinds of implications that moral philosophers have been so concerned, over so many years, to finally conclusively refute the Humean Theory.

But notice that none of these arguments actually turns on making any particular assumptions about what *kind* of psychological state is necessary in order to explain a reason. No matter what kind of psychological state is necessary in order to explain a reason, it is fairly implausible that we are going to be able to expect that everyone, no matter what she is like, will have some psychological state of the requisite kind in order to explain a reason that is supposed to be a reason for everyone. So the Broad Humean Theory of Reasons best captures what lies at the heart of this kind of worry about the universality or objectivity of morality—the kind of worry that the revisionary Humean takes to be conclusive.

Now if the Narrow Humean Theory of Reasons is the most popular *version* of the Broad Humean Theory, it is easy to understand for purely sociological reasons why it would receive so much attention. But what we can expect for sociological reasons is quite different from what we should demand of good philosophy. There are any number of supposed refutations of the Narrow Humean Theory of Reasons in the literature, all for the purpose of setting aside the kinds of skeptical arguments run by Harman, Foot, and Mackie. But it's simply faulty reasoning to think that if an argument you want to rebut needs the premise that *p*, you can rebut it by refuting *p*+, a stronger premise. If we're really concerned about the kinds of skeptical arguments raised by Harman, Foot, and Mackie, we have to be concerned about the more general Broad Humean Theory of Reasons.

3.3 The Classical Argument for the Humean Theory

So why haven't philosophers critical of the skeptical arguments of Harman, Foot, and Mackie been more concerned about this more general view? Are they philosophically

[7] Foot [1975]. Foot, however, subsequently rejected this view. See, for example, her *Natural Goodness* [2001].

[8] Mackie [1977]. The interpretation of Mackie's argument from "queerness" is controversial, however, since there are at least two other good candidates for the kind of argument that Mackie intended to offer. Richard Joyce, however, does unambiguously endorse this argument as the best argument for a moral error-theory, in the process of motivating his moral fictionalism. See Joyce [2002].

lazy? No; a much better explanation is easy to find. The better explanation is that it is widely believed to be common knowledge what the *only motivation* for believing the Broad Humean Theory of Reasons is.[9] And it is an argument which, if it works, also establishes the truth of the Narrow Humean Theory of Reasons. I call it the *Classical Argument* for the Humean Theory.

Elijah Millgram, a critic of the Humean Theory, puts[10] the Classical Argument most succinctly: "How could anything be a reason for action if it could not motivate you to actually *do* something? And what could motivate you to do something, except one of your desires?" Millgram's first rhetorical question states the thesis of *reason internalism* and his second that of the *Humean Theory of Reasons*. If having a reason requires being motivatable, and being motivatable requires having a desire, then having a reason must require having a desire. And that is enough of the Humean Theory of Reasons to motivate the kinds of skepticism just discussed.

A great deal of the abundant literature critical of the Humean Theory of Reasons has focused on rebutting the Classical Argument, and many of the points made there are fairly conclusive. The Classical Argument leaves much to be desired, as a motivation for the Humean Theory of Reasons. But if this is the only motivation for the Broad Humean Theory, then we can straightaway draw two conclusions about the kind of view that the Humean Theory takes about desires. First, they have to be motivating states. And second, they have to be *ubiquitous* motivating states: any action whatsoever has to have one of them in its causal etiology.

These two conclusions set enormous constraints on the kind of shape that the Broad Humean Theory of Reasons might take. If they are sound, then refutations of the Broad Humean Theory of Reasons can take for granted some fairly strong conclusions about what kind of psychological state explains reasons, according to the Humean: not only that they are *desires*, but what desires, in fact, *are*. But I think that if we are genuinely interested in the kind of view that can motivate Harman's, Foot's, and Mackie's kinds of skepticism about the objectivity of morality, then we should cast our nets wider. In particular, I don't think that the Classical Argument gives the best or most interesting argument for the Broad Humean Theory of Reasons. It is the purpose of this paper to offer a better and more general motivation for the Humean Theory, one which doesn't commit that theory to any particular story about what explains the difference between Ronnie and Bradley. It is my purpose to show how *few* assumptions about the Humean Theory of Reasons are necessary in order to motivate it.

[9] Hubin [1999, 31]: "I think what is special about the Humean position on reasons for acting is approximately what most defenders and detractors alike are prone to point to as its attraction . . . What attracts many of us, to the different degrees that we are attracted, to Humeanism is, as many have suggested, a motivational argument."

[10] Millgram [1997, 3]. The classical argument is given in Williams [1981], cited in Bond [1983] and Darwall [1983], and discussed extensively in Korsgaard [1986], Hooker [1987], Millgram [1996], and in many other places. Of these authors, Darwall is the only one who maintains that there are other motivations for the Humean Theory of Reasons.

3.4 The Positive Motivation

It is fairly uncontroversial, as I suggested in section 3.1, that the difference in Ronnie and Bradley's reasons is due to a difference in their psychologies. It is not uncontroversial, of course, *which* difference in their psychologies it is due to. But the central idea behind my motivation for the Humean Theory is to take what we *do* know about Ronnie and Bradley's case, and to put it to work. If there is *any* uniform explanation of all reasons, then maybe what we know about how *some* explanations of reasons work will help to shed light on how *all* explanations of reasons must work. And that is the idea that I will be pushing. There are broad-based theoretical motivations to hope that there might be some common explanation of why there are the reasons that there are—broad motivations to be in search of a uniform explanation of all reasons. If we are after a uniform explanation of all reasons, I will be suggesting, Ronnie and Bradley's case is where we should look.

This may not move you. You may be thinking, 'but maybe there are *two kinds* of reason—one kind that gets explained by psychological states, and one kind that doesn't!' I agree. There *may* be two kinds of reason. But on the face of it, the reason for Ronnie to go to the party and the reason for Ronnie not to murder are both *reasons*—they are both cases of the same general kind of thing. It would be very surprising if these two uses of the word 'reason' turned out to be merely homonyms. So, given that they are both cases of the same kind of thing, it is reasonable to wonder whether there is anything to be said about why they are. And it is this reasonable thing to wonder, I will be suggesting, which will lead to the hypothesis that all reasons are explained in the way that Ronnie's is.

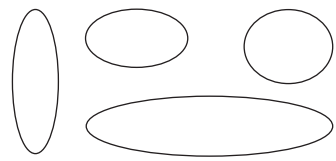
Of course, it doesn't follow from the fact that Ronnie's reason is explained, in part, by his psychology, and the hypothesis that there is a common explanation of all reasons, that psychological features figure in all of these explanations. It could be that the feature of Ronnie's psychology plays a *role* in the explanation of his reason that can be filled by other kinds of thing—for example, by promises or special relationships. And in any case, if we really care about finding a common explanation of all reasons, something must motivate us to pay attention to Ronnie and Bradley's case, in particular. After all, there are *many* cases of reasons, and we might know something about how *many* of them work. Where does the pressure come from to try to generalize Ronnie and Bradley's case to cover others, rather than trying to generalize other cases to cover Ronnie and Bradley's?

This last question is really what this paper is about. My aim is to give a principled motivation for looking to cases like Ronnie and Bradley's. And it will come in two steps. First I'll give a principled motivation from a broad methodological principle for looking to cases of reasons that are *merely agent-relational*, rather than to reasons that are *agent-neutral*, in a sense that may be unfamiliar, but which I will explain. The second, more controversial, step will be to isolate psychology-explained reasons as a better candidate to generalize from than other categories of merely agent-relational reason, such as those deriving from promises or from special relationships. The first

step will occupy the remainder of part 2; I'll offer two arguments for the second in part 3, and another in part 4.

3.5 A Methodological Principle

The argument that if we are looking for a uniform explanation of all reasons, merely agent-relational reasons are the most methodologically promising place for us to look, trades on what I think should be an uncontroversial *methodological principle*. I'll uncover this principle in two stages. First, suppose that you start noticing a lot of shapes like the ones depicted below. These shapes seem to have something interesting in common, and if you investigate, you will be able to find all kinds of interesting things about them. They are, for example, the shape that objects which are actually circular occupy in our visual fields, and so if you are, for example, a painter, it would behoove you to learn more about what they really have distinctively in common that explains why they are *that* shape, rather than some other. It might, after all (indeed, it will), help you to recreate them accurately.



But you'll be going about things all wrong if you start trying to figure out what these shapes distinctively have in common that distinguishes them simply by looking at *them*. It will put you off on all sorts of wild-goose-chases. For example, one of the first things you're likely to notice about your shapes is that they are all round. But what ellipses all have distinctively in common—for the shapes that you are trying to investigate are ellipses—is not simply that they are all round *plus something else*. You won't ever find something that you can add to their being round, to give you the right account of what sets them aside as a distinctive class of shapes. To discover the answer to that, you have to look not only at *ellipses*, but at *foils*—shapes that are like ellipses, but not. In particular, you will want to look at egg-shapes and other non-elliptical ovals. Features that are shared by both ellipses and egg-shapes can be quickly set aside as irrelevant. The Methodological Principle, then, is this:

MP If you want to know what makes *P*'s *P*'s, compare *P*'s to things that are not *P*'s.

I want to take this carefully in order to be perfectly clear how uncontroversial the Methodological Principle should be, because I want to emphasize exactly how natural and forceful my motivation for the Broad Humean Theory of Reasons is. But lest I be accused of belaboring the obvious, the Methodological Principle quickly generalizes once we start paying attention to the case of relations. And here my example will be

slightly contrived. Suppose that having discovered what ellipses have in common[11] you decide to start investigating the *ancestor of* relation, which has always puzzled you. It follows from a generalization of the Methodological Principle that some people are not going to be particularly worth investigating, if you are trying to discover what the common explanation is, of what makes one person the ancestor of another.

Eve, who is the ancestor of everyone (I warned you this would be *slightly* contrived) will not be a particularly good place to start, in investigating the *ancestor of* relation. Since she is the ancestor of everyone, she has no non-descendants to compare to her descendants as foils. And so you will suffer from an embarrassment of riches, if you try to sort through all of the things that all of Eve's descendants have in common, in search of the one that makes them her descendants. Since every human being is one of Eve's descendants (as I stipulated), any feature that every human being shares will become a candidate, and you will have no way of ruling any of these out. So Eve's case gives you no privileged *insight* into the *ancestor of* relation.

If you really want to investigate that relation, the generalization of our Methodological Principle tells us that you need to pay more attention to cases like that of Japheth. Japheth is the ancestor of many people, but he is also not the ancestor of many others. And so we have lots of non-descendants of Japheth to compare to lots of descendants of Japheth. With so many foils, we'll be able to rule out many more potential candidates for what it is that makes Japheth the ancestor of the people who are his descendants. In fact, it is quite likely that there will be *only one* natural candidate for what all of Japheth's descendants have in common but his non-descendants lack: that they are people to whom he stands in the ancestral of the *parent of* relation. So it is quite likely that Japheth's case is going to help you to zero in very quickly on the common explanation of what makes someone the ancestor of someone else. The Generalized Methodological Principle says, then, to pay attention to cases like that of Japheth:

GMP If you want to understand what makes $x_1 \ldots x_n$ stand in relation *R*, compare cases in which $A_1 \ldots A_n$ stand in relation *R* but $B_1, A_2 \ldots A_n$ do not, in which $A_1 \ldots A_n$ stand in relation *R* but $A_1, B_2, A_3 \ldots A_n$ do not, and so on.

Since everyone is a descendant of Eve, Eve's case sets an important *constraint* on a good account of the *ancestor of* relation. That is why it is a relief to check and see that Eve does, in fact, stand in the ancestral of the *parent of* relation to everyone. But by the Generalized Methodological Principle, her case is not the right kind of case to give us any particular *insight* into what makes someone the ancestor of someone else. And that is because it leaves us with no useful foils. It allows us to see things that ancestor dependent pairs have in common, but since it leaves no foils, focusing on this case is like trying to understand ellipses without comparing them to other shapes. It doesn't rule enough out.

[11] They consist in the set of points whose summed distance from each of two fixed points is the same. (This knowledge *will* help you to depict them more accurately, if you really are a painter, because by tying a thread around two pins, you can use this knowledge to trace any ellipse you like with indefinite accuracy.)

3.6 . . . Applied to the Case of Reasons

My *ancestor of* case is, as I noted, slightly contrived. It is highly unlikely, to say the least, that Eve is *really* the ancestor of *everyone*. To be so, she would have to be her own ancestor, which seems rather unlikely to be the case, stipulations aside. So to that extent, the *ancestor of* relation really only approximates the troubles that beset us when we turn our attention to the *reason* relation. For one of the most philosophically salient features of the *reason* relation—and one that we should have fully in view, if we understand the puzzles about the objectivity of morality raised by the Humean Theory—is that there are some reasons that really *are* reasons for everyone, no matter who she is or what she is like. These *universal*, or *agent-neutral*, reasons of morality, about which the Humean Theory of Reasons is supposed to raise so many puzzles, are supposed to be such reasons. Agent-neutral reasons, in the uncontroversial sense, are like the case of Eve, in that they are reasons for everyone.[12] They may place *constraints* on a good theory about the common explanation of reasons, but they *can't* give us any important *insight* into what makes some consideration a reason for someone to do something. For in their case we suffer from an embarrassment of riches. There are too many things that everyone has in common for the case to give us any insight into what distinguishes people for whom *R* is a reason to do *A* from those for whom it is not.

So by the Generalized Methodological Principle, it follows that if you want to know what the common explanation of all reasons is, agent-neutral reasons like the reason to help Katie are not going to be a promising place to start. The only *promising* place to start is with the case of reasons that are *merely agent-relational*: reasons for some people but not for others. Ronnie and Bradley's is such a case. And so Ronnie and Bradley's case is a much more promising place to look, in order to discover what makes reasons reasons than the case of the agent-neutral reason to help Katie, or any of the other moral reasons.

[12] Unfortunately, both the words "universal" and "agent-neutral" turn out to have misleading associations. See my Chapter 2, this volume and Schroeder [2007b], for discussion of the difference between the controversial and uncontroversial senses of "agent-neutral." In essence, in *The Possibility of Altruism* Nagel (although using the terms "objective" and "subjective" at the time) made an uncontroversial distinction between reasons that are reasons for everyone, and reasons that are reasons for only some [Nagel 1970]. But Nagel also adopted the controversial assumption that the only kind of action that a reason can be in favor of, is an action of the form, "promote state of affairs *p*." Only given this highly controversial background assumption does Nagel's uncontroversial distinction, which I am putting to use, succeed at tracking the issues of "agent-relativity" and "agent-neutrality" that have anything to do with the distinction between consequentialism and deontology. The distinction I am making here therefore has nothing directly to do with the existence of agent-centered constraints, of special obligations, or of agent-centered options.

It is also important to distinguish *universal* reasons from *universalizable* reasons. A reason is *universal* if it is a reason for everyone. A reason is *universalizable*, if its existence follows from a general (universal) principle, of the form, "for all *x*, if *x* is in conditions *C*, then there is a reason for *x* to do *A*." So reasons can be universalizable without being universal. See also my 'Cudworth and Normative Explanations' (this volume) for further discussion of this important distinction. For my purposes, getting confused about this is worse than getting confused about whether the distinction has something to do with agent-centered constraints or options, and so I've elected to retain the term "agent-neutral" as the less confusing of these two options.

And that is an interesting result. We might have thought that Humeans are obsessed with cases like that of Ronnie and Bradley because they begin with a pre-theoretic prejudice against reasons like the one to help Katie. After all, Christine Korsgaard has claimed repeatedly that the very idea of a Humean Theory of Reasons *starts* with a special focus on reasons like Ronnie's and a chauvinistic attitude about other intuitive examples of reasons, such as the one to help Katie.[13] But the Generalized Methodological Principle explains why it is natural to be interested in cases like Ronnie and Bradley's. For according to the GMP, we *need* to focus on cases of reasons that are merely agent-relational, in order to see what role the agent-place plays in the three-place *reason* relation: *R* is a reason for *X* to do *A*.

But this observation is still insufficient to justify or even motivate the Broad Humean Theory on the basis of our premises. The observation tells us that *merely agent-relational* reasons are the place that we need to look, in order to see what makes reasons reasons, but Ronnie and Bradley's case is only one *kind* of case of merely agent-relational reasons. The observation explains why the efforts of many philosophers to give explanatory accounts of reasons on the basis of paying special or exclusive attention to moral reasons are straightforwardly methodologically unpromising. But it does not justify paying any more attention to psychology-explained agent-relative reasons than to promise-explained agent-relative reasons, special-relation-explained agent-relative reasons, or any number of others, and that is why the methodological principle only gives us the *first* step in our motivation for the Humean Theory.

Compare: Al promises to meet Rose for lunch at the diner. Andy has made no such promise—he's promised his sick mother to visit her at the hospital. The fact that it's time for lunch is a reason for Al to head to the diner. But it's not a reason for Andy to head to the diner—it's a reason for him to head to the hospital. This difference between Al's and Andy's reasons is explained by their respective *promises*, rather than as a matter of what they like or dislike, want or don't want, care about or not. In another case, Anne is Larry's infant daughter. That is a reason for him to take care of her. But unless you are in Larry's family or a particularly close friend, it isn't a reason for you to take care of Anne. Now, *you* might have all manner of reasons to take care of Anne—she might, for example, have been abandoned by her father. But the fact that she is Larry's daughter is not among *your* reasons to take care of her. Here it is Larry's relationship to his daughter that seems to make for a difference in your reasons.

So examples of merely agent-relational reasons are ubiquitous.[14] Our Methodological Principle tells us to look at what is distinctive of merely agent-relational reasons, in order to understand reasons in general. But that isn't yet enough to close in on the

[13] One such argument is the central line of argument in her [1986]; a distinct and more general argument to this effect is implicit in the opening pages of her [1997b].

[14] Again, to be clear, since what I am after is agent-relational reasons in the uncontroversial sense, what is crucial here is that the reason for Al to go to the diner is not also a reason for Andy *to go to the diner*—not that it is not also a reason for Andy to make sure that Al ends up at the diner. This further feature of Al's reason is highly relevant—but it is not what the uncontroversial sense of "agent-relational" tracks.

Humean idea of focusing on Ronnie and Bradley's case, in which the difference in reasons is due to some *psychological* feature. To do that, we need an argument that Ronnie and Bradley's case gives us a *better* insight into what is distinctive of the agent-place in the reason relation than do Al's case or Larry's case. That is, we need to establish an *asymmetry* thesis. My argument for the Broad Humean Theory of Reasons does not rest on ignoring Al's case and Larry's case, or on taking Ronnie's case more seriously. It rests on establishing this Asymmetry Thesis, to which I turn in sections 3.7–3.10.

3.7 Weak Asymmetry

I'd like to offer three motivations for the Asymmetry Thesis, a weak, a middling, and a strong. The weak motivation motivates a weak version of the Asymmetry Thesis, but rests on less controversial grounds, the middling motivates a middling version of the Asymmetry Thesis and rests on middlingly controversial grounds, and the strong motivation motivates a very strong version of the Asymmetry Thesis, but rests on very controversial grounds. I'll summarize the weak motivation in this section, rehearse the arguments for the middling motivation in sections 3.8–3.10, and end up with the strong motivation in section 3.12; the middling motivation is the one on which I wish to place the most weight for the purposes of this paper, but the broad strategy that I am developing for motivating the Humean Theory can be developed in different ways.

One relevant asymmetry between the case of psychology-explained reasons and other cases of merely agent-relational reasons would be if one of these kinds of reason were a better candidate to generalize in order to explain universal or agent-neutral reasons such as the fact that Katie needs help, which is a reason for anyone to help Katie. According to a common view, it is hopeless to generalize what we know about cases like Ronnie's to cases like that of the reason to help Katie, and that is part of why the Humean Theory of Reasons is hopeless. But I have argued elsewhere that it *is* promising to think that the Humean Theory of Reasons may be able to explain agent-neutral reasons such as the reason to help Katie.[15] There is unfortunately no space to rehearse these arguments here.

There is space, however, to consider why it might be thought unpromising to use cases like those of Al and Larry in order to explain reasons like the reason to help Katie. Al has a reason to meet Rose for lunch because of something that he has *done*—some *promise* that he has made. So one might think about contractualist theories of morality as trying to subsume moral reasons under the case of promises, as in Al's case, in this way. But whatever the promise of contractualism in general, we can only use it to subsume reasons like the one to help Katie under cases like Al's if it is based on *actual* contracts, not merely on *hypothetical* contracts. Al has a reason to meet Rose for lunch because he has *actually* made a promise, not because he *might* have made such

[15] Schroeder [2007a], [2007c].

a promise, if things were different. So only a contractualism based on actual promises could succeed at subsuming moral reasons to cases like Al's. Since that seems unpromising, this seems like an unpromising way to go.

What about cases like Larry's? Could it be that merely agent-relational reasons like Larry's, based on the fact that he is Anne's father, be used to explain reasons like the reason to help Katie? Well, not unless it turns out that everyone is Katie's father. So that doesn't look like a promising view, either. Some authors, however, seem recently to have suggested that being a *fellow human being with* someone is relevantly similar to being the *father of* someone, and that this general relationship, which everyone bears to Katie, can be used to explain reasons in the same kind of way that the fact that Larry is Anne's father can explain agent-relational reasons that Larry has to help Anne.[16] But even supposing this to be true, it would not really be a case of generalizing what we know about Larry's case to all other reasons, because Larry's merely agent-relational reason to help Anne does not derive from the fact that he is a fellow human being with Anne (we all have that reason to help her) but from the fact that *he* is her *father*.

So it is not at all obvious how to generalize other cases of merely agent-relational reasons in a way that would account for the reason to help Katie. It therefore follows that if I am right that Ronnie and Bradley's case *can* plausibly be generalized to account for such reasons, then there is a relevant asymmetry among the obvious cases of merely agent-relational reasons. If we are to look to *any* kind of merely agent-relational reason for insight into the common explanation of all reasons, as the methodological principle suggests that it should be promising to do, then this asymmetry directs us to look to cases like Ronnie and Bradley's. I haven't discharged the antecedent of this argument, here—that requires another paper.[17] But this illustrates one, weak, way in which we might motivate the asymmetry thesis. In the remainder of part 3 I turn to a *middling* way of motivating the asymmetry thesis that we need, on which I wish to place the most weight for the purposes of this paper. And then in part 4 I will use the results of part 3 in order to state a *strong* version of the asymmetry thesis.

3.8 The Standard Model

Recall that the Methodological Principle does not tell us that cases of agent-neutral reasons *don't matter* for an adequate account of reasons. What it tells us is that like Eve's case, they should operate as a *constraint* on a good account, but they are not likely to give us any particular *insight* into the common explanation of all reasons. My first, weak, strategy for motivating the asymmetry thesis had us look at the prospects for each kind of merely agent-relational reason of being used to account for agent-neutral reasons. My second, middling, strategy for establishing the Asymmetry Thesis goes the other way around. It is to show that most merely agent-relational reasons can be

[16] See, for example, Darwall [2009]. [17] Schroeder [2007c].

subsumed to the case of agent-neutral reasons, but psychology-explained reasons like Ronnie's and Bradley's plausibly cannot. If that is right, then we can treat Al's case and Larry's case as setting constraints on an adequate account of reasons, but like Katie's case, not being particularly good sources of insight into that relation. But if it is right, then we *can't* treat Ronnie's case in this way. And that will be my argument that if we want to look for a common explanation of all reasons, psychology-explained reasons like Ronnie's and Bradley's are the first place that we should look. And this is my presumptive argument for the Broad Humean Theory.

So consider the case of Al and Andy. Al promises Rose to meet her for lunch at the diner, and Andy promises his mother to visit her at the hospital. As a result, the fact that it is almost noon is a reason for Al to head to the diner and a reason for Andy to head to the hospital. But plausibly, this difference in Al and Andy's reasons can be traced back to a reason that they have in common—to keep their promises. One such reason is that breaking promises tends to destroy their usefulness. Another is that breaking promises is a breach of trust. Since this is a reason for Al to keep his promises, the fact that he has promised Rose to meet her at the diner for lunch makes heading for the diner at noon necessary for keeping his promises. And since Andy has promised to visit his mother at the hospital, that makes heading to the hospital at noon necessary for *him* to keep *his* promises. So the facts about what promises they have made explain why going *different* places at noon are *ways* for Al and Andy to do the thing that they both have a reason to do—to keep their promises.[18]

It is non-trivial to hold that the difference in Al and Andy's reasons is explained by a further reason that they both share, in this way. Logically speaking, all that we need in order to explain the difference between Al and Andy, is to appeal to the following *conditional*:

Conditional Promise: For all *x* and *a*, if *x* promises to do *a*, then there is a reason for *x* to do *a*.

Logically speaking, no one need have any reasons whatsoever in order for Conditional Promise to be true. But I appealed to something *further* in order to explain Al and Andy's reasons:

Categorical Promise: There is a reason *r* such that for all *x, r* is a reason for *x* to keep her promises.

In this case, it does seem like Categorical Promise is true. I named two such reasons, and likely there are more. And in this case, that seems to be *why* Conditional Promise

[18] Let me immediately head off one source of misunderstanding. When I say that one reason to keep promises is that breaking promises is a breach of trust, I do *not* mean to be suggesting that there is a *further* agent-neutral reason not to breach trust (but not saying what that reason is), and that since breaking promises is a breach of trust, this reason transfers its force to a derivative reason to keep promises. All I am saying is that the fact that breaking promises is a breach of trust is an agent-neutral reason to keep promises. So the explanation that I gave *discharged* the obligation to say *what* the agent-neutral reason from which Al and Andy's reasons derive *is*. But the explanation that I did *not* give *failed* to discharge this obligation—it merely passed it on to the further claim that there is an agent-neutral reason not to breach trust.

is true. So though Al and Andy's reasons differ, that difference can be traced back to an agent-neutral reason. Some philosophers seem to believe, in fact, that *no* conditional like Conditional Promise could ever be true without being backed up with a categorical reason like that in Categorical Promise.[19] But this would be a bold substantive thesis. Logically speaking, Categorical Promise does not follow from Conditional Promise.

Yet the difference between your reason and Larry's can be explained in this same kind of way. Anne is Larry's infant daughter, and that is a reason for Larry to take care of her, but not a reason for you to take care of her. This, it seems, is because the following conditional is true:

Conditional Child: For all *x* and *y*, if *y* is *x*'s infant child, that is a reason for *x* to take care of *y*.

Conditional Child backs up a reason for Larry to take care of Anne, but it doesn't back up a reason for you to take care of her. But in this case, also, it doesn't seem like Conditional Child is true all by itself. Like Conditional Promise, it seems to be backed up by a reason that you and Larry *share*—one to take care of whatever children you *do* have:

Categorical Child: There is a reason *r* such that for all *x, r* is a reason for *x* to take care of whatever children she brings into the world.

Again, it is easy to come up with such reasons. One is that a person's children are moral subjects who cannot provide for themselves, for whom she is causally responsible. This reason seems to back up Larry's reason to take care of Anne, but to avoid backing up the same reason for you to take care of Anne—Anne, after all, is not *your* child.[20]

Cases like these, in which differences in agent-relational reasons are backed up by an agent-neutral reason, follow what I call the *Standard Model* for reason-explanations.[21] The Standard Model is important and interesting, but all that we need to understand about it here is that in a Standard Model explanation, some class of merely agent-relational reasons is collectively subsumed under an agent-neutral reason from which they derive. What I've illustrated here is that merely agent-relative reasons like Al's and like Larry's can be explained in this kind of way, and hence subsumed under the case of agent-neutral reasons. As such, they place *constraints* on a good account of the common explanation of all reasons, but they don't promise to give us any special *insight* into it.

It is natural to think that all cases of merely agent-relational reasons will be like Al's and Larry's cases in this way—that every time some contingent feature of an agent's circumstances plays a role in explaining why something is a reason for *her* to do

[19] I have written about this theory in detail in Chapter 1, this volume.

[20] Again, I do not mean to be saying that there is some more basic agent-neutral reason to take care of moral subjects for whom one is causally responsible. That would not answer the challenge to say what this reason is; it would only put it off. I only mean to be saying that the fact that your children are moral subjects for whom you are causally responsible is a reason for you to take care of them.

[21] See Chapters 1 and 2, this volume.

something, even though it is not a reason for others to do it, it does so by subsuming her case under a more general agent-neutral reason. The theory that all explanations of agent-relational reasons work in this way is the *Standard Model Theory*. According to the Standard Model Theory, though Ronnie's psychological state does play some role in explaining his reason, the role that it plays is a *contingent* one, that can also be played by other kinds of thing. So the possibility of Standard Model explanations is why it doesn't follow from the conjecture that all reasons are explained in fundamentally the same way, and that Ronnie's reason is explained in part by his psychology, that all reasons are in part explained by psychological features. It gives a natural story about how it could be that all reasons really are explained in the same way, and Ronnie's psychological state plays a role in the explanation of his reason, but there are not psychological states in the explanation of every reason. According to the theory, this is because the *role* played by Ronnie's psychology can also be played by other kinds of thing.

But what I'll argue in the next section is that the class of psychology-explained reasons like Ronnie's *can't* be subsumed under agent-neutral reasons in this kind of way. The Standard Model Theory, that is, is false. And that will be the asymmetry that I will argue gives us middling warrant to hold that Ronnie's case is a more promising place to look in order to see what role the *agent*-place plays in the *reason* relation.

3.9 Is There an Agent-Neutral Reason to Promote Your Desires?

To have a Standard Model explanation of reasons like Ronnie's, we need two things. First, we need an action-type *A* such that in every case like Ronnie's, the action the reason is for is a *way* for the agent to do *A*. And second, we need a reason, *R*, that is a reason for anyone to do *A*. It is easy to see how to construct the appropriate *A* and *R* in the paradigmatic cases in which the Standard Model is motivated. What Rachel has a reason to do on both Monday and Thursday is to write about whatever she is thinking about at the time. And the reason for her to do this is that it has been assigned by her poetry professor, the soporific Professor Smith. Because this is a reason for Rachel to write about whatever she is thinking about, it follows that no matter what Rachel is thinking about, she has a reason to write about that.[22]

But unfortunately, it is quite difficult to construct the appropriate *A* and *R* for the full range of cases like Ronnie's. Here I will assume for the sake of argument that there *is* some action such that all actions in favor of which there are psychology-explained reasons are *ways* of doing this action. For the sake of argument, I will assume that this is the action of *doing what you want*. It is unclear, I think, whether any such action-type will do the required work for the Standard Model, but the issues are complicated. I will confine myself to arguing that even if there is some such action *A*, there is no good

[22] See Chapter 1, this volume, for an extended discussion of Rachel's case.

candidate, *R*, for what the agent-neutral reason is to do this thing. If there is not, then the Standard Model Theory is, I think, wrong, and wrong in an interesting way. The way in which it is wrong leaves a relevant *asymmetry* between psychology-explained and other merely agent-relational reasons. And from the preceding considerations, that means that reasons like Ronnie's are the most promising place to look for a unified explanation of all reasons.

This may seem like a silly view. It may seem obvious that there is a reason to do what you want. But we have to be careful how we understand that claim, and consequently we should be suspicious about whether the thought supports the Standard Model in any way. Compare the following:[23]

Easy: For all *x* and *a*, if doing *a* is what *x* wants, then there is a reason *r* for *x* to do *a*.
Mid: For all *x*, there is a reason *r* for *x* to: do what *x* wants.
Hard: There is a reason *r* that is a reason for all *x* to: do what *x* wants.

The problem is that in order to get a Standard Model explanation of the full range of cases like Ronnie's, **Hard** must be true. But it is not at all obvious that **Hard** is true (that is why I called it "**Hard**"). At best, it is **Easy** that is obvious.

Consider the case of Brett. Brett wants to finish his Ph.D. in philosophy. Working on his dissertation on the pragmatics of context-dependence promotes finishing his Ph.D. in philosophy, and so there is a reason for Brett to work on his dissertation on the pragmatics of context-dependence. Moreover, it is easy to see what this reason is. It is that working on his dissertation will enable him to finish his Ph.D. But Brett also wants to become a rock star. Recording a new album with his band will promote this aim. And so it seems that there is a reason for Brett to record a new album with his band. Moreover, it is easy to see what this reason is. It is that recording a new album with his band is necessary in order to get picked up by a label, and hence in order to become a rock star.

Obviously, the reasons for Brett to do these two things are different. Examples like this (at least, enough of them—one for every want) are enough to make **Easy** true. But for **Mid** to be true, there must be a *further* reason for Brett to *do what he wants*, some fact about the world that is both a reason for Brett to work on his dissertation and a reason for him to record a new album with his band. And for **Hard** to be true, this reason, whatever it is, must also be a reason for Ronnie to go to the party, for Vera to practice playing chess, for Christina to buy a new cookbook, for Bill to hike the Appalachian Trail, and so on. What single state of the world could possibly tell in favor of such a rich and diverse class of actions? I don't see what it could be, and no one who believes that there is such a reason has ever given me a good answer as to what they think that it is, either.

[23] Here I bracket the question of whether these claims are sufficient as stated. We're interested in the view that psychological states like desire play a necessary (but not necessarily sufficient) role in the explanation of reasons. If you think some further condition is also required in order to complete this explanation, by all means build it in. This question is orthogonal to the one that I am pursuing here.

The idea I hear most often is also the most unpromising, so let me set it aside, here. The conjecture that I hear most often is that the reason *r* which makes **Hard** true is just the truth of **Hard** itself! How convenient! Unfortunately, also how circular. Even if the truth of **Hard** does satisfy the condition that **Hard**'s existential quantifier governs, it simply can't be the only thing that does. For in order to be such a reason, it must first be true. But in order for it to be true, there must first be such a reason. So it can't be the only one. The fact that I so often hear this hopeless answer seems to me to be evidence that no one does have any good idea of what consideration it could be that makes **Hard** true.

So despite appearances, it should not be at all obvious that there *must* be *some* agent-neutral reason to do what one likes. What should be obvious is that a Standard Model explanation of psychology-explained reasons like Ronnie's owes us something significant. It is committed to holding that there *is* some such reason. And so it should be able to tell us what this reason is. I myself don't know what this reason is. I have no *proof* that there is no good answer as to what it is, but no one, no matter how confident that there *must* be some such reason, has ever given me a satisfactory answer as to what it is. And so I remain suspicious that their convictions that *there is* such a reason arise not from knowing what it is, but because they are in the grip of a theory—the Standard Model Theory. This constitutes my second, middling, motivation for the asymmetry thesis.

3.10 The Argument in Brief

So in sum, this is my argument for the Broad Humean Theory of Reasons, given the middling motivation for the asymmetry thesis:

1 Ronnie's reason is explained by some feature of his psychology.
2 All reasons are, at least at bottom, explained in the same kind of way.
3 From the *Generalized Methodological Principle*, agent-neutral reasons should function as a *constraint* on a good unified explanation of reasons, but they don't give us a promising place to look for how that explanation *works*.
4 From the *Asymmetry Thesis*, all merely agent-relative reasons *other* than the psychology-explained ones can be successfully subsumed under the case of agent-neutral reasons.

C: So psychology-explained reasons like Ronnie's are the *most methodologically promising* place to look for features of how the uniform explanation of all reasons must work.

I don't claim that this argument gives more than a *presumptive* motivation for the Broad Humean Theory of Reasons. All it tells us is that Ronnie and Bradley's case is a *methodologically promising place to look* for an explanation of reasons, *so long as* we aspire for a uniform explanation. But I *do* claim that this argument gives us a *very good* presumptive motivation for the Humean Theory, which is all that I am after.

Premise **1** is weak enough to be uncontroversial—or at least, to create a quite significant cost to rejecting it. Premise **2** is *not* uncontroversial, but it represents an appropriate and reasonable ambition for philosophical theory. Premise **3** is backed by a genuinely uncontroversial methodological principle. And I've argued carefully for premise **4** in sections 3.8 and 3.9—if you think it is false, you're welcome to propose what the action and reason could possibly be that would make a Standard Model explanation of all of the reasons like Ronnie's turn out to work, without raising problems of its own. And if that fails, there is still the weak motivation for the asymmetry thesis from section 3.7. Once we recognize the Methodological Principle and apply it to reasons, we only need *some* relevant asymmetry in order to generate *some* kind of motivation for the Broad Humean Theory of Reasons.

3.11 Revisionist and Conservative Humeanism

Notice that I have *not* claimed that Katie's case, Al's case, Larry's case, and others like them, do not place important *constraints* on an account of reasons. On the contrary, I compared these cases to that of Eve in the *ancestor of* case. Though Eve's case did not in and of itself give us any special insight into the *ancestor of* relation, I claimed that it did place an important constraint on a successful account of that relation. Similarly, I claim that Katie's case, Al's case, and Larry's case place important constraints on a successful account of reasons. I hold that it is a serious mark against any theory of reasons that it fails to account for such reasons.

Distinguish two kinds of Humeanism—*revisionist* and *conservative*. The revisionist Humean is happy to embrace the kinds of skeptical results about the objectivity of morality that I discussed in section 3.2. When the revisionist Humean says that all reasons must be explained by a psychological state just like Ronnie's is, she means that there is no special reason for everyone to help Katie, nor for Al to meet Rose for lunch, and so on. But when the *conservative*, or *sophisticated*, Humean says that all reasons must be explained by a psychological state just like Ronnie's is, he doesn't mean to be denying that there is a reason to help Katie; he is merely making a theoretical claim about that reason's genesis.[24]

The sophisticated Humean's theory may ultimately fail to successfully explain all of the reasons for which he wants to account. If it does so, then he is forced to take a revisionist view. And that can lead, ultimately, to skeptical results about the objectivity of morality. But the motivation that I am offering for the Broad Humean Theory of Reasons is, at least initially, *sophisticated* in outlook. What I am offering is simply a methodological consideration in favor of expecting that Ronnie and Bradley's case should give us a special *insight* into what explains all reasons. And *that*, I would have thought, is all that we need in order to have excellent presumptive motivation

[24] See Schroeder [2007c].

for finding the Broad Humean Theory of Reasons attractive. It is certainly enough to dispel the illusion that the only reason anyone would believe the Humean Theory is because they were committed to the Classical Argument. And that should be enough to dispel the idea that motivation by the Classical Argument can be taken for granted when evaluating the prospects of the Broad Humean Theory of Reasons.

3.12 Coda: *How is* Ronnie's Reason Explained?

One of the principal advantages that I've claimed for my motivation for the Humean Theory of Reasons is that it makes no discriminations among *forms* that the Humean Theory of Reasons might take. It leaves for investigation just *how* the explanation of Ronnie's reason actually works—for example, what kind of psychological state explains it, but also many other questions about how the explanation works. Since we've seen that the Humean Theory cannot accept the Standard Model explanation of Ronnie's reason, and since I've argued in part 3 that this explanation is suspicious anyway, I want to close by offering an alternative way of understanding how Ronnie's reason *does* get explained by his psychology, which leads to an interesting conjecture, which leads to a third, strong, version of the asymmetry thesis, and hence a further, related, argument for the Broad Humean Theory of Reasons.

The fact that there will be dancing at the party tonight is a reason for Ronnie to go there, but not for Bradley to go there. And this is because Ronnie, but not Bradley, desires to dance. For this explanation to be true, *something* like the following has to be the case:[25]

Expl: For all agents x, if R helps to explain why x's doing A promotes p, and p is the object of one of x's desires, then R is a reason for x to do A.

Expl is a generalization under which we can subsume Ronnie's case. In Ronnie's case, the fact that there will be dancing at the party tonight *does* help to explain why going to the party will promote one of Ronnie's desires. For it helps to explain why going to the party will be a way for Ronnie to go dancing, and dancing is something that Ronnie desires to do. But since Bradley doesn't desire to go dancing, it doesn't follow from **Expl** that this is a reason for Bradley to go to the party.

The Standard Model Theory would have it that positing generalizations like **Expl** is not enough to explain Ronnie's reason. For on the Standard Model Theory, as we have seen, **Expl** itself needs to be explained. *Why* is it that **Expl** is true? On the Standard Model Theory, this question must be answered by appealing to a *further* action that there is a reason for everyone to do. But as I've argued, we *can't* successfully do that in this case.

[25] The account given here is the one that I defend in *Slaves of the Passions*, but the details are irrelevant for this point.

But that doesn't mean that **Expl** must be unexplained. Compare **Expl** to another explanatory generalization. We can say that the Bermuda Triangle is a triangle, in part, because it has three sides. This is because the following generalization is true:

Tri: For all *x*, if *x* is a closed plane figure consisting of three straight sides, then *x* is a triangle.

But no one thinks that for **Tri** to be true, there has to be a further shape, over and above triangularity, that is had by everything, and explains why everything has the conditional property postulated by **Tri**. On the contrary, people are likely to think that **Tri** is true simply because it states *what it is* for something to be a triangle. It is because triangularity *consists* in being a closed plane figure consisting of three straight sides, that **Tri** is true.

So I offer **Tri** to the Humean as a model for how the explanation of how Ronnie's reason works, if it does not follow the Standard Model. On this view, a desire helps to explain Ronnie's reason, because there being such a desire is part of *what it is* for Ronnie to have a reason. Like the Standard Model, this is a substantive view about *how* Ronnie's desire helps to explain his reason. But it is an intelligible alternative to the Standard Model. And as such, it suggests the following *alternative* simple argument for the Humean Theory of Reasons, based on what we might call the *Standard-Constitutive Conjecture*:

1 Ronnie's psychology helps to explain his reason.
2 The Standard Model does not successfully account for how it does so.
3 Conjecture: the constitutive model of **Tri** is the only alternative to the Standard Model.

HTR: If so, then being in the kind of psychological state that Ronnie is in must be part of *what it is* to have a reason. So in every case of a reason, there must be some such psychological state.[26]

[26] Special thanks to Mark Murphy, Stephen Darwall, David Copp, Sari Kisilevski, Russ Shafer-Landau, Ralph Wedgwood, Rob Shaver, Gideon Rosen, Gilbert Harman, Michael Morreau, Scott James, and Aaron James, and to audiences at the College Park Conference on Practical Rationality and the second annual Wisconsin Metaethics Workshop.

Part 2

If there were no such facts, and we didn't need to make such claims, Sidgwick, Ross, I, and others would have wasted much of our lives. We have asked what matters, which acts are right or wrong, and what we have reasons to want, and to do. If Naturalism were true, there would be no point in trying to answer such questions. Our consolation would be only that it wouldn't matter that we had wasted much of our lives, since we would have learnt that nothing matters.

[Parfit 2011 volume 2, 367]

4

What Matters About Metaethics?

4.1 Why Parfit's Life Has Not been Wasted

According to Part Six of Derek Parfit's *On What Matters*, some things matter.[1] Indeed, there are normative *truths* to the effect that some things matter, and it matters that there are such truths. Moreover, according to Parfit, these normative truths are cognitive and irreducible. And in addition to mattering that there are normative truths about what matters, Parfit holds that it also matters that these truths are cognitive and irreducible. Indeed this matters so much that Parfit tells us that if there were normative truths, but that these truths were non-cognitive or reducible, then he, Sidgwick, and Ross "would have wasted much of our lives" [OWM2 367].[2]

That it would be a consequence of the thesis either of non-cognitivism or of reductive realism that Parfit would have wasted his life is, of course, no evidence against either thesis; it is perfectly possible even for the most brilliant thinkers to waste their lives. Indeed, as any of the students from my introductory ethics course would be quick to point out, it is very difficult to think clearly and objectively about a question in which to have a large personal stake. My undergraduates readily agree that the steak they have is enough to complicate their thinking about moral vegetarianism; so certainly explosive expressions like 'wasted my life' give Parfit the kind of loaded stake in metaethical questions that should make us cautious of trusting his intuitive verdicts in metaethics. Fortunately, as I will argue in this paper, Parfit has not wasted his life, and he would not have wasted his life, even if it turned out that either non-cognitivism or reductive realism turned out to be true.

[1] Parfit [2011], volume 2. Subsequent references to *On What Matters* in this chapter will be given in-line, with reference to the appropriate volume.

[2] Parfit advises me that since he doesn't think 'truths' even *could* be non-cognitive, he believes this is an infelicitous way of formulating his view. However, as I have argued elsewhere (Schroeder [2010b]), it is safe to assume that if metaethical non-cognitivism is true, then some kind of non-cognitivism about truth must be true as well. (Indeed, I argued in Schroeder [2010a] that truth is itself a much more promising application for expressivism than metaethics is.) So if non-cognitivism is true at all, then it is in fact accurate to say that there are 'non-cognitive truths'. This is one of many examples where it doesn't turn out that a view is incoherent simply because Parfit believes that one of its commitments is false.

In arguing that Parfit has not wasted his life, independently of the answer to any metaethical question, I am, of course, arguing against Parfit's own conception of what makes his life worthwhile. This makes my argument, in a certain way, very presumptuous. Parfit clearly believes that the worthwhileness of (much of)[3] his life turns on the answer to questions in metaethics. But even brilliant thinkers can be wrong, and they are more likely to be wrong both about topics that are relevantly distinct from the topics to which they've applied their greatest brilliance, and when their approach to these topics is colored by a deep sense of a personal stake in them. Still, I admit that it is a bold thesis to claim that someone else's conception of what makes their own life worthwhile is incorrect.

But fortunately, it is no more bold—indeed, it is less bold—than Parfit's own pronouncements to the effect that other philosophers have not understood or believed their own views. For example, about me, Parfit says, "Schroeder's worries seem to show that he does not really accept his own view," on such paltry evidence as that I acknowledged the intuitive force of apparent counterexamples to that view and took steps to explain that intuitive force away [OWM2 361]. It is a pessimistic vision indeed for the possibility of philosophical progress, if it turns out that theorists cannot agree about the intuitive force of examples and offer competing theories about where that force comes from! Whereas Parfit's argumentative strategy in Part Six of *On What Matters* requires showing that everyone who seems to disagree with him either does not have the right concepts to disagree at all, or that they do not really accept their own views (Mackie and Williams apparently fall on the former fork, while I fall on the latter—Nietzsche conveniently slips the forks of the dilemma by going insane), my argumentative strategy only requires establishing that the significant value of Parfit's life has not depended on the answer to central metaethical questions. All I claim, therefore, for my presumptuous argument in this paper, is that I am on better grounds to claim that Parfit is wrong about what makes his life worthwhile than Parfit is to claim that I don't really believe my own philosophical views.

Let me begin, therefore, with my master argument that Parfit has not wasted his life. It goes like this:

(1) *Reasons and Persons* constitutes one of the most important contributions of the last century to making progress in our thinking about substantive normative ethics (*premise*).
(2) Making progress in our thinking about substantive normative ethics is one of the things that matters most (*premise*).
(3) Parfit is the author and creator of *Reasons and Persons* (*premise*).
(4) So Parfit is the author and creator of one of the most important contributions of the last century to one of the things that matters most (from 1, 2, and 3).

[3] I'll be ignoring this qualification from here forward for illustrative purposes.

(5) No life which involves creating one of the most important contributions in a century to one of the things that matters most has been wasted (*premise*).
(6) So Parfit's life has not been wasted (from 4 and 5).

Where could this argument go wrong? It is valid, and has only four premises, one of which is that Parfit is the author of *Reasons and Persons*, which seems difficult to reject. Moreover, the only cause Parfit could have to reject premise 1 would be modesty; indeed the Oxford promotional materials for *On What Matters* describe *Reasons and Persons* as "one of the landmarks of twentieth-century philosophy." Since substantive normative ethics is only one branch of twentieth-century philosophy, and an underappreciated one at that, it is safe to conclude that any contribution to substantive normative ethics that is also a landmark of twentieth-century philosophy full-stop is one of the greatest contributions to substantive normative ethics.

Premise 5 also looks unassailable; surely if any lives are not wasted, it is lives which make epochal contributions to the things that matter most. And yet premise 2 can hardly be said to be a weakness of the argument, either, for it is hard to see why Parfit himself would have spent so much time preoccupied with the attempt to make progress in substantive normative ethics—both of his books are preoccupied with the possibility of such progress—unless he himself agreed that this matters. So I conclude that the argument is sound. Parfit has not wasted his life.

Of course, Parfit may agree with me that his life has not been wasted, for he believes that there are *irreducible*, *cognitive* normative truths about what matters, and he maintains only that his life *would* have been wasted, if it turned out that either non-cognitivism or reductive realism were true. What is at stake isn't *whether* Parfit's life has value, but *what gives it* value—the fact that he has authored one of the most important contributions to one of the things that matters most, or this somehow coupled with the fact that truths about mattering are cognitive and irreducible. Still, how, then, could things go wrong with my argument, if it turned out that there are normative truths, but those truths are either reducible in some way, or require a non-cognitivist interpretation? My argument doesn't say anything about issues metaethical. So where do they come in?

Well, it seems safe to assume that metaethical debates will have no bearing on whether Parfit is indeed the author of *Reasons and Persons*, and so premise 3 looks safe. But there are two possible ways in which one might think that a problem could arise for one of the other premises, on the basis of metaethical views. First, if there can be no such thing as progress in substantive normative ethics, then premise 1 couldn't be true since it says that *Reasons and Persons* was a great contribution to such progress. And second, if nothing at all matters, then it follows that either premise 2 or premise 5 is false. Which is false will depend on whether we interpret the expression 'one of the things that matters most' so that if nothing matters, then everything is among the things that matter most—i.e., not at all. If we so interpret it, then if nothing mattered, premise 2 would be trivially true, but premise 5 would be false, since some lives are

indeed wasted. Whereas if we interpret this expression so that it entails that something actually matters, then premise 2 would clearly be false if nothing mattered. Either way, the view that nothing matters would plausibly make trouble for my argument.

Fortunately for Parfit's concern that whether his life would have been wasted turns on matters metaethical, there seem to be metaethical views with each of these consequences. By the lights of the sort of crude emotivism espoused by a number of the logical positivists in the 1930s, for example, which is clearly a metaethical view, there does not seem to be anything worth calling 'progress' in normative ethics. Indeed many of the logical positivists were of the opinion that there was no properly philosophical discipline of normative metaethical inquiry at all—again, clearly a metaethical view. Similarly, global error theories seem to be committed to the view that nothing really matters, any more than anything is right or wrong, or good or bad. I don't say that if either of these metaethical views turned out to be true, then Parfit's life would indeed have been wasted, because my argument considers only one sufficient condition among, perhaps, very many, for this to be false. But certainly my explanation of why Parfit's life has not been wasted would run into trouble if either of these metaethical theories turned out to be true. So in that respect, metaethics does look like it matters.

However, now we run into yet another problem. For Parfit claims not only that it matters that certain metaethical views are false. He appears to think—indeed, he could have saved hundreds of pages and many hours of his readers' time if he did not—that it matters that *all* metaethical theories other than his own cognitivist non-reductive realism are false. But so far we've only seen that there are certain metaethical views which are committed to rejecting one of the premises of my argument—we've hardly seen that *all but one* metaethical view is committed to rejecting one of the premises of my argument. Yet that seems to be what Parfit must think. How could that be so?

4.2 Conservative Reductive Realism

For concreteness, and because we know from the text that mine is one of the metaethical views which Parfit believes it matters to refute, let's take the case of the sort of conservative, non-analytic, reductive realism that I've defended in previous work. According to this view, some things matter—indeed, there are normative truths about what matters. But this view hypothesizes that there is an interesting question about *what it is* for something to matter—a question that can be answered in non-normative terms. It is no part of this view that we could do away with normative talk and thought about what matters and replace it with non-normative talk and thought. Similarly, it is no part of this view that substantive normative inquiry into what matters is not an autonomous and important domain of genuine inquiry. It is only a theoretical hypothesis about what it is to matter.[4]

[4] In his response to this paper, Parfit [forthcoming] characterizes me, apparently on the basis of the preceding paragraph, as defending 'soft naturalism'. This is the view that "[t]hough all facts are natural, we need

Indeed, it is intended to be a conservative theoretical hypothesis. If any particular hypothesis about what it is to matter turns out to be inconsistent with other particularly indubitable truths, the proponent of this sort of metaethical view sees that as a strong argument against that particular hypothesis. And if every particular hypothesis about what it is to matter turned out to be inconsistent with other particularly indubitable truths, the proponent of this sort of view would cease to advocate it. Nothing about the outlook of this sort of view is intended to undermine or upset ordinary normative ideas; on the contrary, the whole idea is to hold fixed ordinary normative ideas and try to answer some *further* explanatory questions in a way that is particularly theoretically satisfying.

As I have noted, it is part of the conservative outlook underlying the idea that the reducibility of the normative to the non-normative is a potentially fruitful explanatory hypothesis that no particular reductive hypothesis will count as satisfactory, unless it is consistent with independent truths. That at least some things matter, that there can be progress in substantive normative ethics, and that among the things that matter most is such progress, and that lives that make seminal contributions to what matters most are not wasted, are the right sorts of truths to serve as constraints, on this view. The conservative reductive realist is more confident in these truths than she is in the reducibility of the normative. That is what makes her view conservative. But it does not follow from this that she does not believe in the reducibility of the normative after all, as Parfit claims about me. It simply follows that she believes that there is at least one hypothesis about how the normative could reduce to the non-normative that is compatible with all of the most important such independent truths.

Now it may be that the reductive realist has been over-optimistic, and that she is wrong about this. Indeed, there is much that I am inclined to think that I was over-optimistic about in my own first book. (There is always a danger, for ambitious explanatory theories, of falling victim to optimism.) If so, then it may be that on the best available hypothesis about how the normative could reduce to the non-normative, it follows that certain fairly plausible independent normative truths are false, and hence there would be excellent grounds to reject the reducibility thesis. But it is certainly part of the conservative reductivist's *view* that there is an available reductive hypothesis which will *not* predict the falsity of any important independent truths. So

to make, or have strong reasons to make, some irreducibly normative claims" [OWM2 365]. However, I am clearly not a soft naturalist. I do not believe that we need to make or ever have reasons to make irreducibly normative claims. Indeed, I do not even believe that there are such things as irreducibly normative claims. I only believe that there are normative claims, which some people—Parfit among them—erroneously believe to be irreducibly normative. It is part of my view—part of conservative reductive realism—that we can and should make normative claims. It's part of my view that some things matter, and it's part of my view that we couldn't easily dispense with words like 'matter' and still succeed at saying all of the interesting things that we want to say about what matters. But it is no part of my view that we should make irreducibly normative claims, or even that there are such things as irreducibly normative claims for us to make.

if this sort of conservative reductivism were true, then some things would still matter, among them making progress in substantive normative ethics, and such progress would still be possible, as evidenced by, for example, *Reasons and Persons*. The bar is low for a reductive view to be able to explain why Parfit's life has not been a waste; it needn't be consistent with all of the important independent truths; only with those articulated by premises 1, 2, and 5.

It is worth comparing the conservative reductive realist to the flamboyant reductive realist. Whereas the conservative reductive realist is more confident in a range of important independent truths than she is in the reducibility of the normative to the non-normative, and more confident in the reducibility thesis than in any particular hypothesis about how it works, the flamboyant reductive realist is more confident in his reductive hypothesis than in a range of important apparent truths with which it might come into conflict. The conservative reductive realist's attitude toward normative inquiry is that there are other good theoretical questions that are also worth asking. In contrast, the flamboyant reductive realist's attitude is that metaethical problems are so pressing that virtually any plausible answer is worth giving up antecedently compelling normative views, if necessary.

The flamboyant reductive realist may or may not hold that my premises 1, 2, and 5 are compatible with his reductive theory. If he does, then even if his view were true, then Parfit's life would still not be a waste. But there is a natural sense in which the compatibility of premises 1, 2, and 5 with his view is not itself a particularly important part of the flamboyant reductivist's view, for he would be happy to reject these premises if it turned out that he was not able to maintain them. Although this doesn't exactly get us the conclusion that were the flamboyant reductivist's view true, my argument would be unsound, it is not exactly comforting, either. It is therefore understandable why Parfit would want to reject the position of the flamboyant realist, because like the conservative realist, his confidence in truths like premises 1, 2, and 5 is high. It is much less clear, however, why it is important whether the conservative reductivist is wrong.

Just to be perfectly clear about the structure of this point, we may characterize conservative reductive realism as the conjunction of the following four theses:

CRR1: Some things matter, there can be progress in substantive normative ethics, and lives that make seminal contributions to what matters most are not wasted.
CRR2: There is an analysis of what it is to matter that ultimately bottoms out in non-normative terms. This analysis lets us answer explanatory questions that Parfit does not appear to be interested in.
CRR3: If theses (CRR1) and (CRR2) are incompatible, then thesis (CRR2) is false.
CRR4: Theses (CRR1) and (CRR2) are not incompatible.

Because conservative reductive realism is the conjunction of these four theses, in order to observe what implications it has for my argument about the value of Parfit's life, we need to think about what follows if all four of these theses are true. But if all four of these theses are true, then I think it clearly follows, as I've already demonstrated, that

my argument goes through. So it is clear that the value of Parfit's life cannot turn on the question of whether conservative reductive realism is true.

Obviously, Parfit believes that my thesis CRR4 is false. Because he believes this, and because he presumably takes comfort in arguments similar to mine that his life has not been a waste, it is rational for him to hope that my thesis CRR2 is false, and that there is no analysis of claims about what matters that ultimately bottoms out in non-normative terms. But conservative reductive realism is a package view, and there is no rational cause for Parfit to hope that the package turns out to be false.[5]

4.3 The Triviality Objection

Parfit does offer an argument which is presented as an argument against any form of (non-analytic) reductivism. He appears to be quite taken with the argument, as it recurs repeatedly. Moreover, since he devotes six whole pages of *On What Matters* to rehearsing how the argument applies to my view in particular, as a general principle of charitable interpretation, I take it that it is safe to assume that Parfit believes that this argument does, in fact, apply to me, or at least show something instructive about the views that I have defended.[6] The argument is called the 'Triviality Objection', and it is very simple. Parfit begins by defining 'positive' so that if (A) is a generalization of the form, 'When Bx, Dx', where B is a condition spelled out in non-normative terms and D is a normative condition, (A) counts as 'positive' just in case (A) states or implies that when x is B, x also has some other, different, normative property. Similarly, although it plays no direct role in the argument, Parfit defines *substantive* to apply to (A) just in case we might disagree with it, or it might tell us something that we didn't already know [OWM2 343].

[5] In his forthcoming response to this paper, Parfit mistakenly claims that I have said that I am "not really committed to [my] reductive view." This is based on a clear misreading; on the contrary, I am both committed to my reductive view and to the thesis that this view is consistent with the fact that many things matter. The fact that I have the second commitment, which Parfit thinks is an error, does not show that I do not have the first commitment. Indeed, most interesting disagreements among philosophers involve disagreeing about two or more things at the same time. I suspect that a great deal of *On What Matters* could have benefited from greater appreciation of this important fact.

[6] Actually, Parfit goes on to say that the triviality objection applies only to *soft* naturalists [OWM2 344], and do not apply to *hard* naturalists. See note 4 for Parfit's definition of soft naturalism, and my explanation of why I am not a soft naturalist. According to *hard naturalism*, "Since all facts are natural, we don't need to make such irreducibly normative claims. The facts that are stated by such claims could all be restated in non-normative and naturalistic terms." Parfit treats his distinction between hard and soft naturalism as exhaustive, but insofar as I understand this definition, I do not believe that I am a hard naturalist, either. At least, though I do not believe that we need to make any any irreducibly normative claims, that is only because I do not think there is any such thing as irreducibly normative claims to make. I do not accept many of the claims accepted by Sturgeon, Jackson, and Brandt that Parfit goes on to criticize in his discussion of hard naturalism [OWM2 368–377]. What I do believe, is that all normative properties and relations have analyses that ultimately bottom out in non-normative terms. So I suspect that here, as throughout Part Six of *On What Matters*, Parfit is arguing more by consideration of paradigms than by elimination.

With these definitions in hand, Parfit's main presentation of the Triviality Objection considers the example of reductive utilitarianism, but assures us that his argument can be extended to other reductive theses. Since reductive utilitarianism is not, I think, a very plausible view, defending it is not, I think, very interesting for our purposes. Of course, there is a long and venerable tradition in metaethics of arguing against reductivism in general by arguing against straw men and then baldly asserting that one's arguments generalize, but it would not do for us to indulge Parfit in perpetuating this tradition.[7] So since Parfit claims that the same style of argument can be extended to any reductive view, it will be far more instructive for our purposes to consider the general form of the argument. Hence I will assume, in setting out the argument, that we are dealing with an arbitrary reductive view, according to which to be D is just to be B. He calls this thesis (C), and calls the corresponding thesis that when Bx, Dx, (A). He then argues:

(1) (A) is a substantive normative claim, which might state a positive substantive normative fact.
(2) If, impossibly, (C) were true, (A) could not state such a fact. (A) could not be used to imply that, when some act would [be B], this act would have the different property of being [D], since (C) claims that there is no such property. Though (A) and (C) have different meanings, (A) would be only another way of stating the trivial fact that, when some act would [be B], this act would [be B].

Therefore this form of Naturalism is not true. [OWM2 343–344]

I have to confess that Parfit's Triviality Objection is one of the most puzzling arguments I have ever encountered in philosophy. It is true that according to (C), (A) could not be used to imply that when some act would be B, it would have the different property of being D, because according to (C) B and D are the same property. But that is neither here nor there, because premise (1) does not entail that (A) must be able to imply that when some act would be B, it would have the different property of being D. It only entails that when some act would be B, it would have *some* other, different, normative property. This needn't be the property of being D at all. So ignoring the fact that Parfit's

[7] Michael Huemer [2005] takes this tradition to new heights:

> On the face of it, wrongness seems to be a completely different kind of property from, say, weighing 5 pounds. In brief:
>
> 1. Value properties are radically different from natural properties.
> 2. If two things are radically different, then one is not reducible to the other.
> 3. So value properties are not reducible to natural properties.
>
> [. . .] To illustrate, suppose a philosopher proposes that the planet Neptune is Beethoven's Ninth Symphony. I think we can see that that is false, simply by virtue of our concept of Neptune and our concept of symphonies. Neptune is an entirely different kind of thing from Beethoven's Ninth Symphony. No further argument is needed. [94]

second premise gratuitously presupposes that the conclusion of the argument is not only true, but necessarily true, the argument is not even valid.[8]

Moreover, the fact that Parfit seems to treat this argument as if it were valid, by assuming that premise (1) really entails that the 'different normative property' which (A) states or implies must be the property of being D, makes the argument look trivially question-begging. I grant that Parfit is very confident that no reductive theory is true, and that gives him great confidence that for any reductive hypothesis (C), the corresponding statement (A) will state or imply that when x is B, it has the *different* property *of being D*. But what is at issue here is precisely what rational grounds there are for this sort of confidence. And it is very hard to see where any rational grounds for confidence in Parfit's premise (1) are supposed to come from, that do not stem directly from confidence that being D is not the same as being B. And so it is very difficult to see how this argument is supposed to give us any leverage in evaluating whether the reductive hypothesis could be true.

Still, since even if the argument is effectively question-begging, it is not even valid, we can grant Parfit's premise (1) without trouble, so long as attributions of 'D' carry implications that attributions of 'B' do not. If any of these implications are normative, then 'When Bx, Dx' would be positive after all, in Parfit's stipulative sense—even if the reductive thesis is true. In fact, this is a direct consequence of a view for which I've argued in a number of places—namely, that claims about reasons carry *pragmatic* implications about the *weight* of those reasons (which is a normative matter).[9] There is no reason why claims about what would be part of the explanation of why the object of

[8] In his response to this paper, Parfit [forthcoming] suggests that by clarifying how his argument works, we can see that it is clearly valid:

> (1) (A) is a substantive normative claim which might state a positive substantive normative fact.
>
> (2) If, impossibly, (C) were true, (A) could not state such a fact.

Therefore

> (C) is not true.
>
> As Schroeder would agree, this argument is valid. If we knew both that (A) might state such a normative fact, and that if (C) were true (A) could not state such a fact, we could infer that (C) is not true. (Parfit [Forthcoming 2])

Let's again ignore that this argument is made valid by virtue of Parfit's gratuitous inclusion of the presupposition that its conclusion is necessary in premise 2, and assume that what is at issue is whether the argument is valid in some way that non-trivially involves a role for premise 1. Parfit here seems to be suggesting that what does the work in this argument is not the assumption that (A) *does* state a positive substantive normative fact, but only that it *might* do so. But now again we may observe that this argument is not valid (ignoring the illicit presupposition of premise 2), for yet a different reason. Suppose that Derek was in either Hawaii or Alaska last week, but we don't know which. We do know this: if he was in Alaska, then he could not have been in Hawaii. But of course, we don't know where he went. So he *might* have been in Hawaii. From this we cannot infer that he was not in Alaska—only that he *might* not have been in Alaska. What this case illustrates is the general and familiar fact that *modus tollens* is not valid for conditionals with modals in their consequents. So similarly, all we can conclude from the argument if we understand premise 1 in this way is that (C) *might* not be true. But of course, that is where we started—in ignorance of whether (C) is true. So this accomplishes nothing. It is genuinely bewildering to me what this argument is supposed to accomplish.

[9] See especially Schroeder [2007a], chapter 5.

someone's desire would be promoted by her doing something would carry this same pragmatic implication.

In his helpful elaboration of how the Triviality argument works against my view, Parfit contends that if I wish to accept that the 'When Bx, Dx' claim corresponding to my view is positive by his definition of 'positive',

> Schroeder would then face the Lost Property Problem. It is hard to see what this other property could be. And if Schroeder could find some other property that could be the normative property . . . he would have to apply his Naturalism to this other property. The Triviality Objection would then apply to this other claim. This objection would not have been answered. [OWM2 359]

This sounds on the face of it like quite an impressive problem—that it should be both difficult to see what the 'Lost Property' might be, and that even were I to say what it is, we would simply be off on a regress.

Fortunately, however, as I've already noted, it is not difficult to see what other property might be implied by generalizations about reasons, at least according to the views I've already defended in print; it is the property of being a relatively *weighty* reason. And I have in fact already applied my reductivism (unlike Parfit I don't use the term 'Naturalism', which I find unhelpful) to this other property; I've given a reductive account of the weight of reasons in terms of reasons in chapter 7 of *Slaves of the Passions*. Contrary to Parfit, moreover, this does not start the dialectic about the Triviality Objection all over again with the other property, because on my view, there is only one reduction of a normative property or relation in non-normative terms. The 'extra property' that is implied is one that reduces in non-normative terms only *by way of reducing to reasons*. In fact, I've argued elsewhere that *all* promising reductive views should adopt this structure.[10]

Consequently, we may safely reject Parfit's Triviality Objection. It neither provides evidence against conservative reductive realism like that I've defended, nor grounds to think that it matters whether such reductivism is true or false.

4.4 Orogeny of the Mountain

Up to this point in this paper, I've argued that Parfit's life has not been a waste, admitted that the soundness of my argument depends on the falsity of *some* metaethical views, and maintained that it does not depend on the falsity of all alternatives to Parfit's own metaethical view, but only on the falsity of certain, particularly flamboyant, metaethical theses. And I've shown that Parfit's central argument against reductive theories, in particular, is highly problematic.

Fortunately, there is no reason to think that the sort of reductive theory that would be incompatible with one of the assumptions of my argument that Parfit's life has not

[10] See Schroeder [2005b].

been wasted is more likely to be true, or would be more likely to be true, if reductive realism were true, than the sort of reductive theory that would be compatible with those assumptions. Moreover, there are excellent reasons—all of the reasons making the key assumptions of my argument so compelling—to think that a reductive theory that is compatible with those assumptions is much more likely to be true than a reductive theory incompatible with them. In short, among the available reductive theories, some are better than others, being better candidates for the truth. The better reductive theories are the ones that agree about the important claims that my argument assumes or presupposes.

The same distinction, among better and worse theories—a distinction that we can make by appeal to their fit with independently compelling claims—applies to non-cognitivist theories. Just as reductive realists can be flamboyant or conservative, likewise for non-cognitivists. Whereas Carnap and Schlick made flamboyant claims, most contemporary non-cognitivists share a strikingly conservative orientation. Rather than seeking to derive stunning or unintuitive consequences, they aim to preserve all of the important claims—normative and otherwise—that Parfit emphasizes are so important, and to go on to ask a set of further, explanatory, questions. It's possible to be interested in these further explanatory questions because you find it puzzling *whether* there are any normative truths. But it's also possible to simply be curious about *how* there are normative truths, and find non-cognitivism a promising approach for providing a particularly satisfying answer.

Like the distinction among reductive realist views, there are excellent grounds—grounds provided by a lot of independently compelling truths—to hold that conservative non-cognitivist views are much more likely to be true than flamboyant ones. Holding this does not require holding that conservative non-cognitivist views will be able to bear all of the fruits which they promise—like the reductive realist, the conservative non-cognitivist may be over-optimistic about the resources of her view. Indeed, at times conservative non-cognitivism has largely consisted of optimism.

But even if we are pessimistic about the conservative non-cognitivist's aspirations for success in her conservative ambitions, that's not quite the same as it *mattering* that she fails. We should distinguish *predictions* that conservative non-cognitivism will fail from Parfit's apparent *hope* that it will, and similarly for conservative reductivism. It hasn't been my aim in this paper to defend either reductivism or non-cognitivism. It has instead been my aim to lower the stakes of the discussion so that we can evaluate these theories in reasonable and objective ways, treating them as what they are—theories. Certainly they may be false. But if they turn out not to be, everything will still be okay, so long as some things really matter and moral progress really is possible. And I think we should all have pretty high confidence that *if* any reductive or non-cognitivist theory is true, it is one that is not inconsistent with the fact that many things matter.

The fact that some reductive theses are better than others should look familiar, for readers of parts two, three, and five of *On What Matters*. For in parts two and three

Parfit argues that some Kantian views are better than others, and in parts three and five he argues that some Contractualist views are better than others. Together with his view that some Consequentialist views are better than others, this leads to the result that any Kantians, Contractualists, and Consequentialists who share Parfit's confidence in the data that motivate discriminating these better versions of these views from the worse versions, have much to agree about. Rather than arguing against Kantianism or Contractualism *as such*, Parfit argues only against the versions of Kantianism and Contractualism which fall astray of this core set of data. What turns out to be important, for the Parfit of the core chapters of *On What Matters*, is not which of Kantianism, Contractualism, and Consequentialism is true, but the core theses which their best versions share.

Another similar phenomenon arises in one of the most surprising twists of the entire book, on page 467, just a few pages into his discussion of the metaphysical objections to non-reductive normative realism, when Parfit launches into a criticism of *actualism* and defense of *possibilism*. This is not a defense of the view in ethics known as 'possibilism', but of the thesis from the metaphysics of modality that there are possibilia which don't actually exist. Since possibilism is typically seen as a particularly ontologically extravagant thesis, this is hardly the move one expects in a chapter whose ostensible purpose is to persuade us that Parfit's view is metaphysically innocuous. Yet including Appendix J, Parfit spends a full forty pages attempting to defend this view, even going so far (don't be surprised) as to allege that "though Plantinga claims to be an Actualist, that is not really true" [OWM2 739].

One leaves the appendix with the distinct impression that the thesis that Parfit cares about is simply not the thesis over which participants in the debate in the literature on the metaphysics of modality between actualism and possibilism disagree. Rather, what Parfit seems to think is important, and the reason why he seems to think that Plantinga is really a closet possibilist, is merely that there be a way for us to talk about the different options that an agent could take in a choice situation—something that actualists and possibilists might make sense of in different ways.

In much of part six of *On What Matters*, I'm tempted to suspect that something very similar has happened, for metaethical inquiry in general. There is something important that Parfit is concerned about, and there are real views in metaethics that are inconsistent with the results that he needs—views on which, in particular, my argument that Parfit's life has not been wasted is unsound. But I'm inclined to think that the important issue about which Parfit cares is not quite the same as the issues that have been pursued in contemporary metaethical inquiry under the headings of reduction or non-cognitivism. Rather, if what Parfit cares about is right, then though many metaethical views are indeed false, there is still a striking range of what I've called *conservative* metaethical theories—views which share a relatively common picture of the data, but offer competing explanations of it. Though all but one of these views are false, which one turns out to be true would not affect whether Parfit's life has been wasted, and will have no consequences for Parfit's arguments in the core chapters of *On What Matters*.

Like the convergence between Kantian, Consequentialist, and Contractualist approaches to normative theory, the conservative approaches to metaethics which I've been discussing here share a common conception of some of the data. But unlike them, I don't believe that they could merely be complementary paths toward the same truth (although contrast Gibbard [2003]). Rather, they are loosely like different orogenies for the same mountain—different theories about where it came from.

If what you are primarily interested in, like Parfit, is how to get to the top of the mountain, then you may not care where the mountain came from. And if most of the people you talk to who do care where it came from are mostly concerned to try to convince you that that since they can't understand where it came from, it must really be a flat plain, or that since they can't understand how you could have gotten so high, you must not be climbing the same peak as anyone else, you are not likely to find orogeny very worthwhile. But it doesn't follow that the mountain has no history. Even fellow climbers can pause, every once in a while, to admire the sweeping vistas, to rest up for the next leg of the journey, and to ponder whether this mountain was formed by subduction, volcanic action, or in some other way. It is true that many contributions to metaethics are like the orogenist telling Parfit that there is no mountain, or that everyone has her own mountain. But at its best and most interesting, metaethical inquiry needn't be like that at all. It has room for many questions which can be pursued with an open mind even by mountaineers who share Parfit's quest for the peak.

5

Supervenience Arguments Under Relaxed Assumptions

With Johannes Schmitt

5.1 Introduction

When it comes to evaluating reductive hypotheses in metaphysics, supervenience arguments are the tools of the trade. Jaegwon Kim and Frank Jackson have argued, respectively, that strong and global supervenience are sufficient for reduction, and others have argued that supervenience theses stand in need of the kind of explanation that reductive hypotheses are particularly suited to provide. Simon Blackburn's arguments about what he claims are the specifically problematic features of the supervenience of the moral on the natural have also been influential. But most discussions of these arguments have proceeded under the strong and restrictive assumptions of the S5 modal logic. In this paper we aim to remedy that defect, by illustrating in an accessible way what happens to these arguments under relaxed assumptions and why.

The occasion is recent work by Ralph Wedgwood [2007], who seeks to defend non-reductive accounts of moral and mental properties together with strong supervenience, but to evade both the arguments of Kim and Jackson and the explanatory challenge by accepting only the weaker, B, modal logic. In addition to drawing general lessons about what happens to supervenience arguments under relaxed assumptions, our goal is therefore to shed some light on both the virtues and costs of Wedgwood's proposal [2000].

5.1.1 Some Hasty Background

To understand Wedgwood's attempt to explain supervenience within a non-reductivist framework, we need a few tools from modal logic. It is often assumed that necessity and possibility are governed by what is known as the S5 modal logic, which amounts to

making the following three assumptions (the letters we use for these assumptions are typical in modal logic[1]):

(K) $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$
(T) $\Box\varphi \supset \varphi$
(E) $\Diamond\varphi \supset \Box\Diamond\varphi$

In the context of the other axioms, axiom (E) is in turn equivalent to, and it is useful to break it down into, the conjunction of (B) and (4):

(B) $\Diamond\Box\varphi \supset \varphi$
(4) $\Box\varphi \supset \Box\Box\varphi$

The intriguing move that Wedgwood makes is to give up on the idea that all of these assumptions are true, and in particular that (4) is mistaken.

The standard semantics for modal logic allows us to interpret what these axioms amount to, by postulating a two-place 'accessibility' relation among worlds, R, such that '$\Diamond$P' is true when evaluated at some world, w, just in case there is some world, w*, such that R(w,w*), and 'P' is true when evaluated at w*. Given this interpretation, axiom (E) says that the relation R is Euclidean, which means that if any world is related by it to each of two other worlds, then they are related by it to each other. Similarly, axiom (B) says that R is symmetric, and axiom (4) says that it is transitive. (E) and (T) together imply that R is an equivalence relation, and it is since R is an equivalence relation under the S5 axioms, that assuming the S5 axioms allows us to ignore R altogether, and talk simply of 'possible worlds', rather than of possible worlds that are 'accessible' from some particular world.

Just to give a picture, which will be useful in a moment, axiom (E) says that whenever the diagram in figure 5.1 obtains, the situation in figure 5.2 obtains, as well. To coin a term, it says that *absolutely* possible worlds are possible relative to each other. While axiom (4) says that whenever the situation in figure 5.2 obtains, the situation in figure 5.1 obtains, as well. Given our terminology, it says that *relatively* possible worlds are absolutely possible.[2]

The relationship between these two diagrams is crucially important, because both strong and global supervenience can be and have been formulated in different ways in the literature—ways that correspond to each of the diagrams.

[1] For example, in Hughes and Cresswell [1996].

[2] Point of clarification: in our diagrams, we will follow the convention that the absence of an arrow is the absence of information about accessibility, rather than positive information about non-accessibility—so fig. 1 is neutral about whether w_2 is accessible from w_1 and conversely, as well as about whether w is accessible from each of w_1 and w_2, and about whether any world is accessible from itself. To indicate non-accessibility, we will use crossed arrows (as in fig. 6).

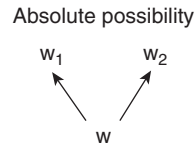


Figure 5.1

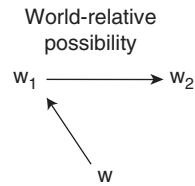


Figure 5.2

5.2 Supervenience Theses

Supervenience theses share the feature of claiming that any difference of one kind (an 'A-difference', in the *supervening category*) must be matched by a difference of a certain other kind (a 'B-difference', in the *subvening category*). They differ with respect to what counts as an A-difference and a B-difference, and with respect to what goes into the 'must'.

Supervenience theses can be formulated either in natural language, using terms like 'necessarily' and 'possibly', or they can be framed in terms of quantifiers over possible worlds. Characterizations of supervenience claims in terms of possible worlds should not be taken to be committed to the view that 'necessarily' and 'possibly' are object-language quantifiers over possible worlds, or even that they are to be understood in terms of metalanguage quantifiers over possible worlds. Even if 'necessarily' and 'possibly' are primitives, we take it that a suitable notion of 'possible worlds' can be constructed, so that we can use equivalences between natural language claims and possible-worlds talk in order to try to introduce clarity to matters that would otherwise be difficult to sort out.[3] In this and immediately following sections, because we are interested in the logical relationships between supervenience theses, we will follow the convention of framing these theses in terms of quantifiers over possible worlds, as these make these logical relationships clearer. In part 6 we will return to consider what it would take to state equivalent versions of these theses in ordinary English, on the grounds that it is easier to test the intuitive plausibility of sentences formulated in ordinary language, which is what will interest us in part 6.

[3] The issues surrounding quantifying over merely possibly possible worlds, under the assumption that axiom (4) is false, are complex; we propose to finesse those issues here by ignoring them.

5.2.1 Strong Supervenience

Take strong supervenience first. According to strong supervenience, any two items at any two worlds which share all of their B-properties must also share all of their A-properties. That is, any worlds w_1 and w_2 satisfy the Strong Supervenience Condition,[4] or SSC, for short:

SS $\forall w_1 \forall w_2 (SSC(w_1,w_2))$

This is all well and good, if we make the assumptions about necessity and possibility that are codified in S5. But if we deny (E), then it is important to know whether and how the world-quantifiers in SS are restricted. According to one way of precisifying SS, the *absolute* formulation, or SS^{abs}, both world quantifiers range over worlds that are accessible from w—the world with respect to which SS is being evaluated. Whereas according to a second way of precisifying SS, the world-relative formulation, or SS^{WR}, the w_1 quantifier ranges over worlds accessible from w, while the w_2 quantifier ranges over worlds that are accessible from w_1.

SS^{abs} $\forall w_1{:}R(w,w_1) \forall w_2{:}R(w,w_2)(SSC(w_1,w_2))$
SS^{WR} $\forall w_1{:}R(w,w_1) \forall w_2{:}R(w_1,w_2)(SSC(w_1,w_2))$

So SS^{abs} is true at w just in case for every situation like that depicted in figure 5.1, w_1 and w_2 satisfy SSC. And SS^{WR} is true at w just in case for every situation like that depicted in figure 5.2, w_1 and w_2 satisfy SSC. So it follows that since according to (E) every figure 5.1 situation is a figure 5.2 situation, whatever goes for figure 5.2 situations goes for figure 5.1 situations as well, and hence SS^{WR} together with (E) entails SS^{abs}. Similarly, since according to (4) every figure 5.2 situation is a figure 5.1 situation, what goes for figure 5.1 situations goes for figure 5.2 situations as well, and hence SS^{abs} together with (4) entails SS^{WR}. That is why so long as we assume S5, it doesn't matter which formulation we use, but as soon as we drop S5, it matters a great deal—one could be true and the other false.

Moreover, one supervenience claim, SS^{abs}, say, could be true at one world in a given model without being true at another world of that same model. To see this, suppose we have a B-model consisting of three worlds with accessibility-relations as shown in figure 5.3 by the arrows. Suppose, moreover, that there is only one subvenient property, G, and only one supervenient property, F, and that in all three worlds w, w_1, and w_2, a is

[4] $SSC(w_1,w_2) \equiv \forall x{\in}D(w_1)\ \forall y{\in}D(w_2)(\text{B-indiscern}(x,y){\rightarrow}\text{A-indiscern}(x,y))$, where '$D(w_1)$' picks out the domain of world w_1—the class of objects existing at that world, and '$\text{B-indiscern}(x,y)$' is an abbreviated way of saying that *x* and *y* share all of their B-properties—i.e., that for each property in B, *x* has it if and only if *y* does ($\forall G_{\in B}(Gx{\equiv}Gy)$)—and similarly for '$\text{A-indiscern}(x,y)$'. Wedgwood formulates strong supervenience somewhat differently, the main difference being that his formulation requires the assumption that the set of the B-properties is closed under Boolean operations; the formulation used in this note does not require this assumption. In the main text we'll ignore the precise characterization of the Strong Supervenience Condition, the better to make clear how the (E) and (4) assumptions affect the relationship among different kinds of supervenience.

Figure 5.3

the only object that has G. If a has F in w and w_1 but ~F in w_2, SS^{abs} will be true at w, but will fail at w_1.

In the following, if we talk about a supervenience claim like SS^{WR} being true *tout court*, this is to be understood as an elliptical way of saying that the claim is true at the *actual* world.

5.2.2 *Global Supervenience*

Like strong supervenience, global supervenience admits of both absolute and world-relative formulations. Intuitively, global supervenience says that any two worlds which are the same in the distribution of their B-properties are the same in the distribution of their A-properties. Or for short, that any two worlds satisfy the Global Supervenience Condition[5] (GSC, for short):

GS $\forall w_1 \forall w_2 (GSC(w_1,w_2))$

Again, we can distinguish between whether both quantifiers range over worlds accessible from w, or whether the w_2 quantifier ranges over worlds accessible from w_1, yielding absolute and world-relative versions of global supervenience—GS^{abs} and GS^{WR}:

GS^{abs} $\forall w_1{:}R(w,w_1) \forall w_2{:}R(w,w_2)(GSC(w_1,w_2))$
GS^{WR} $\forall w_1{:}R(w,w_1) \forall w_2{:}R(w_1,w_2)(GSC(w_1,w_2))$

By the same reasoning as before, these bear the same logical relationship to one another. Since GS^{abs} quantifies over figure 5.1 situations and GS^{WR} quantifies over figure 5.2 situations, the fact that (E) says that every figure 5.1 situation is a figure 5.2 situation means that given (E), what goes for figure 5.2 situations goes for figure 5.1 situations, and hence that GS^{WR}, together with (E), entails GS^{abs}. Similarly, the fact that (4) says that every figure 5.2 situation is a figure 5.1 situation means that given (4), what goes for figure 5.1 situations goes for figure 5.2 situations, and hence that GS^{abs}, along with (4), entails GS^{WR}.

In this section, we have noted that both strong and global supervenience theses admit of both absolute and world-relative formulations, and that their relationship depends on some of the substantive assumptions that are encapsulated in the S5 modal logic. In particular, the absolute versions of these theses entail their world-relative versions only under the assumption of (4), and the world-relative versions entail the

[5] $GSC(w_1,w_2) \equiv (\text{B-indiscern}(w_1,w_2) \rightarrow \text{A-indiscern}(w_1,w_2))$, where 'B-indiscern($w_1$,$w_2$)' is an abbreviated way of saying that w_1 and w_2 have the same distribution of B-properties. There are different ways of precisifying what it takes for two worlds to have the 'same distribution' of B-properties; the same point goes for each of these, and we won't worry about such details, here.

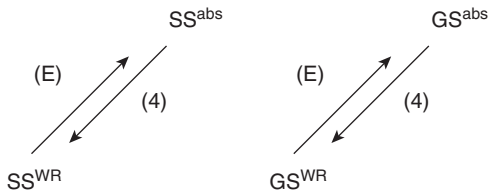


Figure 5.4

absolute versions only under the assumption of (E). Our progress so far is encapsulated in figure 5.4.

5.2.3 The Relationship Between Strong and Global Supervenience

It is often argued or assumed that strong supervenience entails global supervenience, and has sometimes been claimed that global supervenience necessitates some corresponding version of strong supervenience, even though strong supervenience does not follow from global supervenience in every model.[6] In this section we briefly survey the former of these two claims, and how it is affected by relaxing our modal logic.

The idea that strong supervenience entails global supervenience is simple. Global supervenience requires that any two worlds satisfy the global supervenience condition, which says, intuitively, that they differ in their distribution of A properties only if they differ in their distribution of B properties. Strong supervenience says that any two worlds satisfy the strong supervenience condition, which says, intuitively, that any item in the first world differs from any item in the second world in some A property only if they differ in some B property as well. So the argument that strong supervenience entails global supervenience is simple: take two worlds which differ in the distribution of their A properties. Intuitively, there must be something which has some A property in one world but lacks it in the other.[7] If so, then the assumption that these two worlds satisfy the strong supervenience condition tells us that the item which has the A property in one world but not the other must differ in some B property between the two worlds. And intuitively, that suffices to make the two worlds differ in the distribution of their B properties. So on the assumption that any pair of worlds satisfies the strong supervenience condition, we can show that they also satisfy the global supervenience condition.

[6] Kim [1984] originally argued that strong supervenience did follow from global supervenience in every model; countermodels were subsequently given by Hellman [1985] and others, as discussed in Kim [1987]. In an enlightening paper, Paull and Sider [1992] argue that in interesting cases, these countermodels violate combinatorial constraints on the space of possible worlds, and hence strong supervenience may indeed follow from global supervenience in *intended* models. We won't wade into these questions here.

[7] One possible complication with the argument may arise if A includes relations as well as one-place properties. At a minimum, this makes a more careful formulation of the argument necessary.

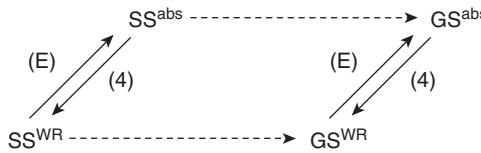


Figure 5.5

We don't mean to endorse this reasoning, here; just to remind readers of it. Making good on this reasoning requires making good on a precise way of understanding global supervenience's talk of the 'same distribution' of properties in each of two worlds. What we want to observe, is that insofar as the reasoning is sound, it gets us the conclusion that each version of strong supervenience entails its *corresponding* version of global supervenience, and that the reasoning is equally sound for the connection between SS^{abs} and GS^{abs}, as for the connection between SS^{WR} and GS^{WR}. This is because the soundness of the reasoning turns only on whether the strong supervenience condition entails the global supervenience condition, once the global supervenience condition is formulated more precisely. We supplement our picture with dashed arrows to indicate our caution about these entailments (figure 5.5).

5.3 Fancy Supervenience Arguments from the Literature

Both strong and global supervenience have been argued to entail a certain kind of *reductivism*: that if one set of properties supervenes on another, then the first set can be *reduced* in some way to the second. It is best to divide these arguments into a more logically straightforward first step, which argues that supervenience of some kind entails a certain kind of necessary equivalence, and a more philosophically loaded second step, which involves the assumption that necessary equivalences of this kind are a sufficient condition for identity.

So, for example, Jaegwon Kim [1984] has argued that if the set of A properties strongly supervenes on the set of B properties, it follows that for each A property, we can construct a property in the Boolean closure of the B properties with which it is necessarily coextensive. This is the logically straightforward first stage of his argument. Then he assumes that properties can be no more fine-grained than necessary coextensiveness, and hence that the condition that he has established is sufficient for property identity. Hence, he concludes that if the A properties strongly supervene on the B properties, each A property must be identical with some property in the Boolean closure of the B properties, and hence that every A property can be analyzed in terms of the B properties, along with Boolean operations. We'll look at this argument and how it fares under our generalized assumptions in a moment.

On the global supervenience side, Frank Jackson [1998] has argued that if the set of A properties globally supervenes on the set of B properties, then for each atomic A proposition, to the effect that some particular thing has one or another of the A properties, it is possible to construct a complex B proposition to which it is necessarily equivalent. This is the logically straightforward part of his argument. The actual dialectic of Jackson's next step is somewhat complicated, because he doesn't directly assume that necessary equivalence is sufficient for proposition identity (though he does think this). What he does instead is to claim that the same argument goes through '*mutatis mutandis*' for properties, to assume that necessary coextensiveness is sufficient for property identity, and to claim that he has established that every A property can be analyzed in terms of the B properties.[8]

It is not clear what Jackson means by saying that his argument goes through *mutatis mutandis* for properties, given that in the text, he has assumed global supervenience but not strong supervenience, and unless Jackson is making the controversial assumption that global supervenience entails strong supervenience (which certainly does not follow given the assumptions we have been making so far), his argument does *not* go through *mutatis mutandis*. Nevertheless, a more careful exponent of the argument from global supervenience to reduction might proceed more directly, and simply assume (what Jackson thinks anyway) that necessary equivalence is sufficient for proposition identity, and hence infer that every A proposition is analyzable as a complex B proposition, staking her claim to have established a reduction on this thesis about the identity of propositions, as opposed to on a claim about the identity of properties.

In the next section we will show that Kim and Jackson's arguments work only for the absolute versions of strong and global supervenience, SS^{abs} and GS^{abs}. This is the payoff of Wedgwood's rejection of (E); by accepting world-relative but not absolute versions of both strong supervenience and global supervenience, his aim is to keep what is true about supervenience without engendering its commitment to the necessary equivalences which Kim and Jackson argue lead to reduction.

5.3.1 Strong Supervenience Arguments Under Relaxed Assumptions

Take Kim's argument first. Assume that the set of A properties absolutely strongly supervenes on the set of B properties, at world w. Our first step is to create a partition of possible individuals, by establishing a set of mutually exclusive B-maximal properties, $B^{\star}$. A B-maximal property is a property with the feature that no two individuals with that property can differ in any of their B properties. It is a sufficient procedure to construct such a set, to take the set of all conjunctions which contain, for each property in B, either it or its negation, yielding a set of B-maximal properties. The set may be

[8] "The same line of argument can be applied *mutatis mutandis* to ethical and descriptive predicates and open sentences: for any ethical predicate there is a purely descriptive one that is necessarily co-extensive with it." Jackson [1998, 123].

reduced to a set of *mutually exclusive* B-maximal properties by dropping one of each pair of members which fail to be mutually exclusive.[9]

To construct a property in the Boolean closure of the B properties that is necessarily coextensive with an arbitrary A property, F, what we do is to partition the members of B* into those which are co-possible with F and those which are not co-possible with F. The disjunction of the former, $\vee\{G \in B^*:\text{co-possible}(G,F)\}$, is the property that we need. Because it is a disjunction of conjunctions of properties in B and their negations, it is in the Boolean closure of B. It is necessarily sufficient for F, because each of its disjuncts is necessarily sufficient for F. To see why that is so, suppose that some property in B*, say, G, is co-possible with F. So there is a world possible relative to w, where some individual is G and F. Absolute strong supervenience then tells us that any individual *at any world possible relative to w* that is not F cannot be G, because it must differ in some B property, and by the assumption that G is in B*, two things that are G cannot differ in any B property. That is, for any individual at any world possible relative to w, if it is G, then it is F. So G is sufficient for F. Wedgwood calls such sufficiency claims *specific supervenience facts*, and we'll return to discuss them, later. Our disjunction is also necessary for F, because the properties in B* form a partition of possible individuals, and so any individual at any world that is F must have some property in B*—which is consequently co-possible with F, and hence in our disjunction. So $\vee\{G \in B^*:\text{co-possible}(G,F)\}$ is in the Boolean closure of B, and necessarily coextensive with F, and a similar construction goes for any other property in A.

The reasoning in this argument is impeccable, but it relies on the absolute formulation of strong supervenience at the crucial place that I've indicated with italics. The easiest way to see that the world-relative version of strong supervenience does not by itself get us Kim's conclusion is to observe that if (E) is false, then F and G may be co-possible because they are both instantiated by a at w_1, possible relative to w, but at w_2, also possible relative to w but not relative to w_1, b is G but not F (figure 5.6). If this is so, then it may be necessary *at* w_1 that everything that is G is F, without being necessary that everything that is G is F.

In fact, the scenario just envisaged is fully compatible with the assumption of (B), provided that G is uninstantiated at w, as figure 5.7 illustrates, and we'll have occasion to comment further on, later.

Figure 5.7 is a (B) model in which the Kim argument fails, and the right-hand model is a (4) model in which it fails. Importantly for Wedgwood, the compatibility with (B) requires that G is uninstantiated at w, as figure 5.7 illustrates, and we'll have occasion to comment further on, later.

So far, the moral is this: absolute, but not world-relative, strong supervenience entails the constructability thesis for properties: that for each supervening property,

[9] The reasoning here will not rely on the members of B* being mutually exclusive, but this will come into play in one of the proofs in the appendix.

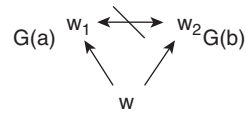


Figure 5.6

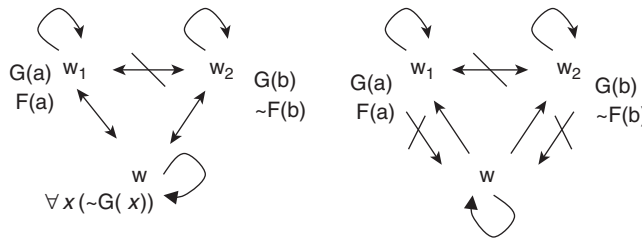


Figure 5.7

there is a property in the Boolean closure of the subvening properties to which it is necessarily coextensive. So if (E) is false, either because (B) is false or because (4) is, then at least one sort of strong supervenience thesis could be true, even though there are no properties in the Boolean closure of the B properties that are necessarily coextensive with any of the A properties.

5.3.2 Global Supervenience Arguments Under Relaxed Assumptions

A similar result goes for the argument from global supervenience to the constructability thesis for propositions. The way that argument works is very similar; instead of constructing a partition of possible individuals by constructing B-maximal properties, we start by constructing a partition of possible worlds by constructing a set, B†, of B-maximal world-descriptions—propositions which describe the entire world in each and every B detail. Global supervenience says, intuitively, that two worlds can differ in the truth of some A proposition only if they differ in which member of B is true at them.

Jackson's construction procedure then asks us to look, for an arbitrary A proposition, P, at each world, w_1, at which P is true. Because the elements of B† form a partition, some member of B†, Q, is true at w_1. And so absolute global supervenience tells us that at every world possible relative to w, if Q is true, then P is true. Now take the disjunction of each such Q, for each world at which P is true: $\vee\{Q \in B\dagger\text{:co-possible}(Q,P)\}$. This disjunctive proposition is sufficient for P, because each of its disjuncts are, and necessary for P, because it includes a disjunct for every world at which P is true.

Again, the reasoning in this argument is impeccable, but it relies crucially on the assumption of the absolute version of global supervenience. The easiest way to see that the world-relative version of global supervenience does not get us to Jackson's conclusion is to observe that if (E) is false, then it is compatible with world-relative global

supervenience that there may be possible worlds w_1 and w_2, such that P and Q are both true at w_1 and Q and ~P are true at w_2 (figure 5.8).

If this is so, then Q⊃P may be necessary *at* w_1, without being necessary. In fact, as before, this scenario is also compatible with the either (B) or (4), though compatibility with (B) requires that Q is false at w (figure 5.9).

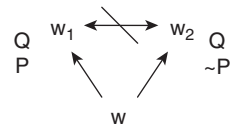


Figure 5.8

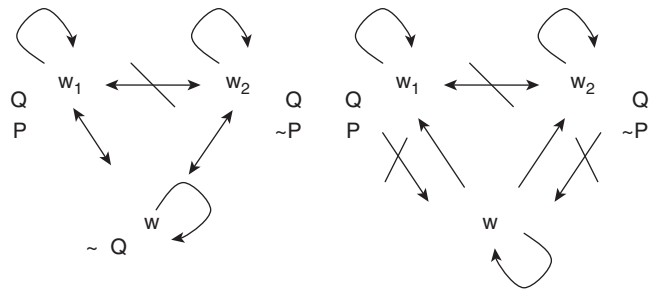


Figure 5.9

So absolute, but not world-relative, global supervenience allows for the constructability thesis for propositions: that for each supervening proposition, there is a proposition in the Boolean closure of the subvening propositions with which it is necessarily equivalent. We can summarize our progress with figure 5.10.

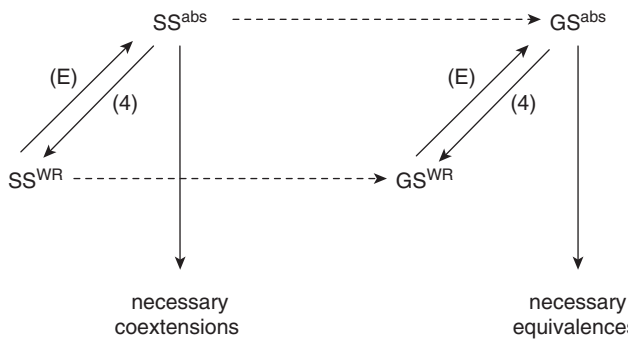


Figure 5.10

5.4 Specific Supervenience Facts

So what, then, follows from the world-relative versions of strong supervenience and global supervenience? In the case of world-relative strong supervenience, we can derive, along with the assumption of (T), what Wedgwood calls 'specific supervenience facts'. A specific supervenience fact is a fact of the form: ∀w:R(@,w)(∀x(Gx⊃Fx)), where F is a predicate for some A property and G is a predicate for some property in the Boolean closure of the B properties. Equivalently, without explicit quantification over worlds, a specific supervenience fact has the form, □∀x(Gx⊃Fx). Such a fact specifies a modally *sufficient* condition for an A-property in terms of B-properties, without being a *necessary* condition for it.

To derive a specific supervenience fact for some supervening property F, find some member G of B*, our class of mutually exclusive B-maximal properties, such that F and G are *actually* co-instantiated, i.e., we have F(a)&G(a) for some individual a in the domain of the actual world, @. (T), recall, implies that R(@,@), so from (T) and SS^{WR}, we can derive ∀w:R(@,w)(SSC(@,w)) by universal instantiation. Now to show that □∀x(Gx⊃Fx), we'll let w be an arbitrary world possible with respect to @, and let x be an arbitrary member of w's domain, and show that G(x)⊃F(x). By the B-maximality of G, our assumption that G(a) at @, and that G(x) at w, it follows that a and x are B-indiscernible. (If they weren't, then they would differ in some B-property, and hence G wouldn't be B-maximal.) So it follows from SSC(@,w) (which we derive from ∀w:R(@,w)(SSC(@,w)) by universal instantiation) and the rule of detachment that a and x are also A-indiscernible. But since we assumed that F(a) at @, it then follows that F(x) at w. Hence we have shown G(x)⊃F(x) at w, and since w and x were arbitrary, we get □∀x(Gx⊃Fx).

As we noted one paragraph back, this is a one-way entailment. World-relative strong supervenience allows us to construct one such one-way entailment for every B-maximal property with which F is actually co-instantiated. Each of these one-way entailments is a *specific supervenience fact*. We can also take the disjunction of all of the antecedents of these specific supervenience facts, ∨{G ∈ B*:actually co-instantiated(G,F)}. This property, too, is sufficient for F. But we cannot get a *two*-way entailment, because nothing forces this disjunction to be *necessary* for F. Any B-maximal property that is not actually instantiated will not be actually co-instantiated with either F or ~F, and this argument works only for B-maximal properties that are actually co-instantiated with F. This should be no surprise; we already saw in section 5.3.1 that an uninstantiated B-maximal property may be both co-possible with F and co-possible with ~F. So this disjunction gives us a property that is necessarily coextensive with F just in case every B-maximal property is actually instantiated. Hence the constructability thesis for properties holds at all and only worlds where every B-maximal property is instantiated.

In keeping with the tight correspondence between how things work for strong and global supervenience, world-relative global supervenience allows for an analogue of specific supervenience facts, something that we call *specific actuality entailments*. To

construct such an entailment, first locate the B-maximal world description, Q, which describes the actual world in every detail. Next choose some A-proposition, P, which is actually true. Since we have R(@,@) (which follows from (T)), GS^{WR} can be simplified to ∀w:R(@,w)(GSC(@,w)), which, given the truth of Q and P at @, implies that at every world possible relative to the actual world at which Q is true, P is true as well. But that means that Q⊃P is necessary at the actual world.

Again, this is a one-way entailment. We can construct similar one-way entailments for each other A-proposition that is actually true, but we can't use world-relative global supervenience to construct any other one-way entailments whose consequents are P, because we don't have any guarantee that what is necessary at other possible worlds will actually be necessary.

Specific actuality entailments are the least controversial sort of evidence for some kind of supervenience of the normative on the non-normative. I am as certain as I could possibly be of anything that it is impossible for the world to be exactly as it is in every non-normative respect, but different normatively in that the fact that my mother is my mother is a reason to torture her. I may not know exactly *how* the facts about the non-normative world guarantee that things could not be different in this normative way without some normative difference, but I am absolutely certain that they could not. So this is an example of a specific actuality entailment. It makes sense that a thesis like this one should be the most obvious thesis involved with supervenience claims, because as we've seen so far, it is the weakest sort of view that one could take (figure 5.11).

5.4.1 Explanatory Arguments Based on Supervenience

Kim and Jackson's arguments aspire to show that supervenience *entails* reduction. As we noted, each of these arguments divides into two components. The first is an

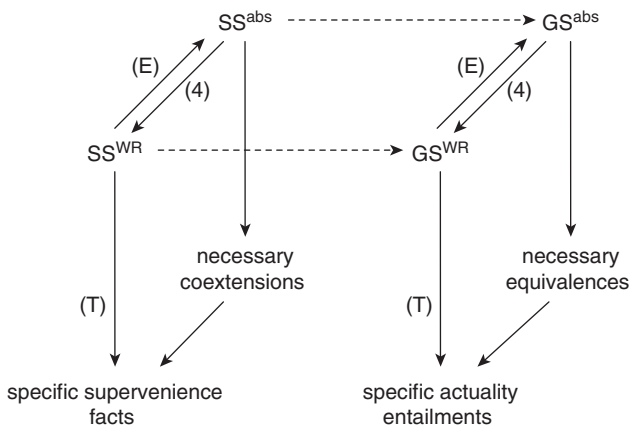


Figure 5.11

uncontroversial *constructability* thesis, to the effect that A properties or propositions have necessarily coextensive or necessarily equivalent counterparts from the Boolean closure of the B properties or propositions. The second is a controversial assumption that necessary coextensiveness is a sufficient condition for property identity, or that necessary equivalence is a sufficient condition for proposition identity. So these direct arguments from supervenience to reduction rely on the very strong assumptions, respectively, that properties and propositions are very coarsely individuated. A more cautious set of arguments from supervenience to reduction does not rely on these assumptions.

The key idea of these more cautious arguments is that even if supervenience theses do not *entail* reduction, they still leave something to be *explained*. The idea is that since supervenience theses tell us not just that certain things *don't happen* to occur, but that it is *impossible* for them to occur, they aren't the kind of thing that can just happen by coincidence. If the A properties are genuinely distinct from and irreducible to the B properties, then why should it be *impossible* for something to have such-and-such a B-maximal property and not have so-and-so an A property?

This reasoning is reinforced by consideration of natural recombination principles. Toy blocks can be square or round, and red or blue. Even if there are not any red square blocks, there could have been—all that it would take, would be for there to be a block that 'recombines' the square feature that some blue blocks have with the red feature that some round blocks have. Since these are distinct properties that blocks can have, they can be recombined in any combinatorially possible way. Every combination is possible, unless something prevents it. So if A properties are distinct from B-properties, what prevents them from being recombined in any possible way? What makes some recombinations not just non-actual, but *impossible*? Supervenience, it seems, leaves something to be explained.

Reductive hypotheses can explain supervenience. If A properties are reducible to B properties, then we can explain why they supervene on the B properties in the same way as we would explain why the property of being a red square supervenes on the properties of being square and of being red. Since supervenience hypotheses stand in need of explanation, and reductive hypotheses seem particularly well-suited to explain supervenience, there is therefore an explanatory argument from supervenience to reduction. Unlike the Kim–Jackson style arguments, it is only a defeasible argument. But unlike them, it does not rely on controversial assumptions to the effect that properties or propositions are very coarse-grained.[10]

Moreover, explanatory arguments do not require the absolute versions of strong or global supervenience in order to get started. In fact, they don't even require the full-blooded assumption of *any* supervenience thesis to get started. Any necessary connection between A properties and B properties will do, including specific

[10] See, for example, Schroeder [2005b].

supervenience facts or specific actuality entailments. Even these weaker theses leave something to be explained—something that reductive hypotheses can offer an explanation of.

5.4.2 Where We Are

So far we've seen that there are two main kinds of arguments connecting supervenience to reducibility. The first kind, due to Kim and Jackson, seeks to show that supervenience *entails* reducibility, and works by first establishing a constructability result to the effect that every A property (proposition) has a necessarily coextensive (equivalent) counterpart from the Boolean closure of the B properties (propositions). The second, philosophical, step of such arguments, proceeds by assuming that necessary coextensiveness (equivalence) is a sufficient condition for property (proposition) identity. The advantage of this kind of argument is that if it works, its conclusion is strong: that supervenience entails reducibility. Its weakness is the strong philosophical premise that it requires: that properties (propositions) are so coarse-grained.

The other kind of argument connecting supervenience to reduction is the explanatory argument. According to the explanatory argument, supervenience need not entail reducibility, but it does leave something important to be explained, and reductive hypotheses are the right sort of thing to fill this explanatory gap. The advantage of this kind of argument is that it does not rely on the assumption that properties or propositions are so coarse-grained. Its disadvantage is that it is less conclusive, without some exhaustive classification of the possible strategies for explaining the modal facts that are involved with supervenience.

The good news for non-reductivists who would like to accept supervenience theses is that when we relax the restrictive assumptions of the S5 modal logic, we also strengthen what is required in order for Kim and Jackson's constructability arguments to work. As we saw, those arguments work only for the absolute versions of strong and global supervenience, and not for the world-relative versions, except in S5, in which the two versions are equivalent. This means that relaxing assumption (E) of the S5 modal logic is necessary but not sufficient in order to escape these constructability arguments.[11] If you want to accept supervenience theses but avoid these arguments, you need to be careful to accept only the world-relative supervenience theses, and to deny the absolute versions.

[11] To be more precise, relaxing assumption (E) is necessary but not sufficient for evading the first, constructability, step of the Kim–Jackson style arguments. It is not necessary for evading those arguments *tout court*, because it is possible to evade them at the second, philosophical step, at which they assume that necessary coextensive (equivalent) properties (propositions) are identical.
Billy Dunaway has pointed out to us (in discussion) that the claims (i) through (iv) are jointly unsatisfiable

(i) ~(E)
(ii) (B)
(iii) necessary coextensiveness is necessary and sufficient for property identity
(iv) if properties P and Q are identical (at a world), they are necessarily identical at that world.

Consequently, even once we relax assumption (E), much more work remains to be done, for the non-reductivist who would like to accept supervenience but escape the Kim–Jackson constructability arguments. Such a theorist must defend her choice to accept only the world-relative supervenience theses, and deny their absolute versions. For what we've seen is that even if we relax our modal logic, everything will still hang on *which* formulation of supervenience is correct. In the remainder of this paper, we will look at Ralph Wedgwood's proposal for how to exploit the relaxed assumptions involved with denying assumption (4), in order to answer the explanatory argument. Like the Kim–Jackson style of argument, we will see that there is good news for non-reductivists who want to accept supervenience. But also like the Kim–Jackson argument, we will see that everything hangs on *which* formulation of supervenience is correct.

Consequently we will close the paper in section 5.6 by evaluating which supervenience theses are more compelling—the absolute or the world-relative versions. Peculiarly, though this distinction does so much work in his arguments, Wedgwood never considers this question explicitly, and in fact never even distinguishes between the two formulations of strong supervenience. We will argue, however, both that the world-relative theses are hard to formulate clearly and unambiguously in natural language, making it implausible that they are what we find pre-theoretically compelling,

The proof (by reduction) is very straightforward: We start with constructing a model in which (E) fails (the square brackets indicate which propositions are true at the worlds and the arrows the indicate accessibility-relation):

$$\mathbf{w}_1\,[\mathrm{p}] \leftarrow \mathbf{w}_2\,[\Diamond \mathrm{p}, \sim \mathrm{p}] \rightarrow \mathbf{w}_3\,[\sim \Diamond \mathrm{p}]$$

Here, $\Diamond$p is true at w_2 because w_2 'sees' w_1, but $\Box\Diamond$p is not, because w_2 also sees w_3; hence we have the antecedent of (E) but not its consequent. Assumption (ii) tells us that accessibility-relations must hold symmetrically across the model:

$$\mathbf{w}_1\,[\mathrm{p}] \rightleftarrows \mathbf{w}_2\,[\Diamond \mathrm{p}, \sim \mathrm{p}] \rightleftarrows \mathbf{w}_3\,[\sim \Diamond \mathrm{p}]$$

Assuming the sufficiency direction of (iii) at w_3, the property $\lambda x.\ p$ is identical with the property $\lambda x.\ x \neq x$ at w_3. Moreover, this identity is *necessary* (by (iv)). Since w_2 is accessible from w_3, the identity must hold in w_2, too. So $\lambda x.\ p$ and $\lambda x.\ x \neq x$ are identical at w_2. By (iv) they are *necessarily* identical. Now, w_1 is accessible from w_2, so the identity must hold in w_1, too. But since p is true at w_1, the extension of p at w_1 is non-empty, assuming that w_1 has a non-empty domain (which we now assume by stipulation). Hence by (iii) again, the two properties cannot be identical at w_1. Contradiction!

We think that this proof is *prima facie* good evidence that one cannot reject S5 (while accepting B) without giving up one of (iii) and (iv). There are some possible complications with the proof (it assumes unrestricted comprehension for properties within a second-order framework) but we think they may not be decisive. It seems to us that denying (iv) is mind-boggling, so we conclude that (iii) has to go (notice that in Dunaway's proof both directions of (iii)—the sufficiency and the necessity—are appealed to).

This basic consequence should really be no surprise; in any ~(4)-model, properties can be necessarily coextensive without being necessarily necessarily coextensive, and necessarily necessarily coextensive without being necessarily necessarily necessarily coextensive, and so on. So intuitively, it is not plausible that necessary coextensiveness is sufficient for property identity after all; a better candidate is *hyper*necessary coextensiveness, where 'hypernecessarily *p*' is defined as the limit of '$\Box p$', '$\Box\Box p$', '$\Box\Box\Box p$'.... But then the constructability argument would need to establish properties that are hypernecessarily equivalent, not just necessarily equivalent.

and that even once we reject S5, denying the absolute versions of the supervenience theses is still particularly unintuitive.

5.5 Explaining Supervenience Given Blackburn's Point

The explanatory argument challenges the non-reductivist to explain the necessary facts that are involved with the claim of supervenience. Ralph Wedgwood's attempt to do so consists in three basic parts, which we'll explain in the next three sections, respectively. The first part of the explanation is to explain *some* of the necessities involved with supervenience by appeal to claims about essence. It is an important fact about supervenience, however, emphasized in Simon Blackburn's arguments about supervenience, that some necessities involved with supervenience cannot be explained in this way, and this leads to the second part of Wedgwood's account, which is to explain the remaining necessities emphasized in Blackburn's point by appeal to contingent explananda—a resourceful move made available by giving up on (4). Finally, the third part of Wedgwood's account confronts what he takes to be a remaining problem arising from global supervenience. Our goal in these three sections will be not only to explain the nature of Wedgwood's account, but to illustrate how it hangs on the choice between SS^{abs} and SS^{WR}, as well as on the choice between GS^{abs} and GS^{WR}, in addition to hanging on the denial of (4).

The first part of Wedgwood's account of supervenience is to propose that the normative supervenes on the non-normative because it is in the *nature* of normative properties to satisfy strong supervenience. Just as it is in the essence of squares to have four sides, it is in the essence of the normative to supervene. This view adopts the way of thinking, familiar from thinkers like Kit Fine [1994], that essence is prior to, and explanatory of, modality, and things are possible because they are compatible with essence (though this becomes slightly more complicated at the next step of Wedgwood's account).

Wedgwood is not the first to suppose that it is part of the nature of the normative to supervene, though he is the first to articulate this thought so clearly in the framework of thinking about essence. It faces an obvious problem, of which he is keenly aware, and it is that the normative cannot supervene, without supervening in some *particular way*. As we observed in section 5.4.1, given (either) version of strong supervenience, we can derive *specific supervenience facts*, which tell us *how* the normative is supervening. For example, nothing about the bare fact of strong supervenience tells us whether the property of being good is compatible or not with, say, the property of causing much more deep unhappiness than it prevents. But if strong supervenience is true, then these properties *may* be incompatible—indeed, there are guaranteed to be some properties that are incompatible, if strong supervenience is true.

Simon Blackburn emphasized this important fact about strong supervenience in three different versions of an argument against moral realism that he presented in

his [1973], [1984], and [1985]. We won't rehearse all of the parts of that problematic argument here, but a central feature of the argument was Blackburn's point that *bare* supervenience is impossible—in order for there to be strong supervenience, there must be some specific supervenience facts or other, even though no particular specific supervenience facts are required for strong supervenience. (Blackburn himself tried to argue that there are no specific supervenience facts of the relevant kind, and used this to argue that strong supervenience should be replaced with *weak* supervenience, which he in turn argued that only non-realists are in a position to do.)

The relevant point for Wedgwood's argument, however, is that it is not enough, to say that it is in the essence of normative properties to supervene, in order to explain all of the metaphysical impossibilities associated with normative properties, if strong supervenience is true. You could, of course, say that it is in the nature of normative properties to be co-instantiated with particular non-normative properties, but that just looks like the reductivist's explanation of supervenience. So Wedgwood needs a different kind of explanation of specific supervenience facts, and that brings us to the second part of his account.

5.5.1 Contingent Explananda for Modal Explanans

Wedgwood's explanation of specific supervenience facts is simple, and takes advantage of giving up (4). His thought is that if it is in the nature of A properties to strongly supervene on B properties, and some particular A property, F, happens to be actually co-instantiated with some particular B-maximal property, G, then it will have to be impossible for anything to be G but not F—otherwise the A properties wouldn't supervene on the B properties after all. So the idea is, having first explained strong supervenience by appeal to a claim about essence, to use strong supervenience, together with *contingent* facts, to explain specific supervenience facts. (Even though it is necessary that something is, if G, F, it is only contingent that anything is both F and G.)

It is easier to see how Wedgwood's explanation works, to think about it in terms of a picture of the space, not only of possibility, but of *hyper*possibility. Let the space of hyperpossibility include not only the possible worlds, but the possibly possible worlds, and the possibly possibly possible worlds, and so on—out to the limit. This space is represented by the model structure for a (B) model of modal logic. Rather than thinking about the space of possibility as constrained only by essence, Wedgwood thinks of the space of *hyper*possibility as constrained only by essence. Anything that is compatible with essence is hyperpossible—including any way of recombining A properties with B properties, because since the essence of the A properties only constrains them to strongly supervene, from Blackburn's point it follows that they might still supervene in any of a number of ways. Consequently, each of these ways of supervening is hyperpossible.

But facts about essence don't just constrain which *worlds* there are in the space of hyperpossibility; they also place primitive constraints on which worlds are possible relative to which others. For example, even though it is hyperpossible for F to

be co-instantiated with the B-maximal property G, and hyperpossible for ~F to be co-instantiated with G, neither of these two worlds is possible relative to the other. At the closest, they are each possible relative to a third world where G is uninstantiated, as in figure 5.6. Consequently, what is possible in this picture depends on *where we are* in the space of hyperpossibility. So this is what Wedgwood's appeal to contingent facts does—it isolates where we are in the space of hyperpossibility, in order to isolate which of the many hyperpossibilities are really, genuinely, possible.

The non-reductivist at whom the explanatory argument is directed accepts supervenience theses and consequently metaphysical impossibilities, without having any explanation for why those things are genuinely impossible. Insofar as they have no explanation, they appear to be brute. But the problem is that metaphysical impossibilities don't seem like the kind of thing that *can* be brute. In Wedgwood's picture, however, the metaphysical impossibilities posited by specific supervenience facts *do* have an explanation—they are explained by contingent facts. And then the explanation stops. But that seems like a much more acceptable place for the explanation to stop—after all, contingent facts *do* seem like the kind of thing that can simply be brute. Ultimately, things simply have to be one way or the other.

So far, so good—the second step of Wedgwood's answer to the explanatory argument exploits the assumption that (4) is false, by working with a model in which not everything that is hyperpossible is genuinely possible. Like the earlier ways of avoiding the constructability results from Kim and Jackson, therefore, it works by exploiting a relaxation in the assumptions built into the S5 modal logic. But that's not all it needs; as we'll see in the next section, Wedgwood's answer to the explanatory argument also shares with the attempts to avoid the constructability results, the feature that it trades on *which* versions of strong and global supervenience we assume. This is best illustrated by introducing the third part of Wedgwood's account.

5.5.2 The Problem of Global Supervenience

The third part of Wedgwood's account is needed to address the problem that if GS^{abs} is true, then there will be further necessary truths which still remain unexplained. The problem goes like this: let Q be a B-maximal world description, as in section 5.3.2. That is, Q is a proposition which describes the world in every B detail, so that necessarily, any two worlds in both of which Q is true are B-indistinguishable. And let P be any A proposition whatsoever. Since GS^{abs} says that worlds which are B-indistinguishable must be A-indistinguishable as well, it follows as a special case that if G is true at two worlds, then P must be either true at both or false at both (assuming bivalence). That is, one of the following two theses will be true:

1 $\Box(Q \supset P)$
2 $\Box(Q \supset \sim P)$

Moreover, this would be a problem for Wedgwood because his explanation of specific supervenience facts does not generalize to explain either of these facts. Wedgwood's explanation of specific supervenience facts, in the second part of his account, can only

explain necessities that follow from supervenience facts, together with contingent facts. But even holding fixed the truth of GS^{abs}—assuming that it might be true in virtue of the essence of normative properties—no contingent fact suffices to explain *which* of 1 and 2 is true, unless Q happens to be true at the actual world. Unfortunately, however, if Q&P is merely *possible*, that suffices, together with GS^{abs}, to mean that 1 has to be true. Similarly, if Q&~P is merely possible, that suffices, together with GS^{abs}, to mean that 2 has to be true. So to explain whichever of 1 or 2 is true, Wedgwood would need to explain either why Q&P is impossible, or why Q&~P is impossible. But nothing about supervenience or about contingent facts suffices to say which of these is impossible—just that one of them *has* to be impossible. This is the revenge of the problem raised by Blackburn's point—to be true, supervenience theses require that one or another of some stronger, more specific necessities also be true, but they don't specify which one needs to be true.

The third part of Wedgwood's account is therefore to reject GS^{abs}. Distinguishing GS^{abs} from GS^{WR}, he endorses only the latter, and rejects the former. Given GS^{WR}, we can construct only *one* necessity of the form $\Box(Q \supset P)$—the one where Q and P are true at the actual world. This is the specific actuality entailment that we constructed in section 5.4.1. But this *particular* necessity *can* be explained by Wedgwood's method from the second part of his account: it is explained on the basis of the contingent fact that Q&P (or that Q&~P—whichever the case may be).

Given that Wedgwood endorses the entailment of global supervenience by strong supervenience, this means that he is committed to denying SS^{abs} as well. Wedgwood is unfortunately less than perspicuous on this point, and fails to distinguish between the absolute and world-relative versions of strong supervenience, as he does for the two versions of global supervenience. The only version of strong supervenience that he even discusses is SS^{WR}. But as we saw in section 5.2.2, if SS^{WR} entails GS^{WR}, then SS^{abs} entails GS^{abs} as well.

So Wedgwood's response to the explanatory argument, just like the responses to the constructability arguments considered in section 5.4.2, requires not only weakening our modal logic (the response to the constructability argument required giving up (E) and Wedgwood's account requires giving up (4)), but also requires the assumption that the absolute formulations of both strong and global supervenience are false. So if he wants to endorse supervenience at all (as he does), he needs to endorse only the world-relative versions of each. In the final part of the paper, we'll now evaluate whether it is plausible to reject the absolute formulations of strong and global supervenience, even in a relaxed modal logic. We'll be arguing that this is still a highly counterintuitive way to go—*even if* (4) is false.

5.6 Absolute Supervenience Theses are Independently Intuitively Compelling

As we have just seen, the viability of Wedgwood's account requires a specific package of views on strong and global supervenience. It requires him to affirm the relative versions

of strong and global supervenience while denying their absolute versions. In this section we will argue that denying the absolute versions is very implausible. We can test for these intuitions by formulating the different versions (and their negations) in natural language and inquiring which of the formulations 'sound' intuitively true and also, on a comparative scale, whether some versions seem more plausible than their counterparts.

We will be arguing not only that it is unintuitive to deny the absolute version of SS (and GS) *per se*, but that it is especially odd to deny them while at the same time affirming their world-relative counterparts. This can be made particularly vivid by considering some of the cases in which supervenience is often invoked, such as the supervenience of the moral on the non-moral or natural. In fact, it is hard to see why any reason[12] to deny the absolute version would not also give us (similar) reasons to deny the relative versions. What is it that sets the two theses apart? We believe that similar points will go for both strong and global supervenience, but here we will confine our remarks to strong supervenience.

Earlier in this paper we gave formulations of the two versions involving quantification over possible worlds. Doing so made it easier to see the logical relationships among supervenience theses. But now we will translate these formulations of supervenience into equivalent natural language sentences that involve modal adverbs like *necessarily* and modal complementizers like *it could have been (the case) that*.[13] Since these expressions are more natural ways of talking than expressions like 'in every possible world', the idea is that framing supervenience using them will make it easier to think about the intuitive plausibility of supervenience theses and their negations.

Another standard way of formulating SS^{WR} and SS^{abs} which can be found in the literature[14] better lends itself to natural-language formulations. It looks like this:

(i) $\Box\forall x\forall F{\in}A[(F(x) \rightarrow \exists G{\in}B^*(G(x)\&\Box\forall y(G(y) \rightarrow F(y))))]$
(ii) $\forall F{\in}A\forall G{\in}B^*[\Diamond\exists x(F(x)\&G(x)) \rightarrow \Box\forall y(G(y) \rightarrow F(y))]$[15]

For the proof of the equivalence of SS^{abs} and (ii), see Appendix A.1. The relationship between SS^{WR} and (i) is slightly more complicated, but it turns out that SS^{WR} holds at all worlds in a model just in case (i) holds at all worlds in that model; we prove this in Appendix A.2. Since Wedgwood's explanation of supervenience in terms of essence

[12] Any reason *other* than the reason that your theory requires you to hold one version while denying the other.

[13] We choose this more complicated way of expressing alethic modality because we think that 'it is possible that p' mostly expresses epistemic possibility.

[14] See, for example, Kim [1984, 65], where Kim formulates world-relative strong supervenience in this way. He does not call it 'world-relative', of course, because he does not distinguish it from absolute strong supervenience.

[15] Note that in (ii) the property-quantifiers have to take scope over the whole conditional—which means that they don't occur within the scope of any modal operator. While (i) requires the assumption that all A-and B-maximal properties exist in all possible worlds (regardless of whether they are instantiated in those worlds), (ii) requires only the (weaker) assumption that all A- and B-maximal properties exist at the *actual* world. This creates a complication we'll need to gloss over, here.

requires it to hold at all worlds in the model (that is, in all hyperpossible worlds), we take it that SS^{WR} and (i) are interchangeable for his purposes.

The advantage of working with (i) and (ii), rather than only with SS^{abs} and SS^{WR}, is that they make it easier for us to look at something closer to natural-language versions of supervenience, starting with (i) (world-relative strong supervenience):

(I) Necessarily, if anything has one of the A properties, then there is a B-maximal property which it has, and which necessarily implies having that A property.

It might seem harder to translate (ii) into English. In fact, most formulations of absolute strong supervenience that can be found in the literature seem to be in a language that allows quantification over possible worlds—similar to our earlier formulation (SS^{abs}).[16]

Here is how one might try to render (ii) in something approximating ordinary English:

(II) For any A property and any B-maximal property, if they could have been co-instantiated, then necessarily, anything that has that B-maximal property has that A property.

Despite doing our best to formulate both SS^{WR} and SS^{abs} in comprehensible, natural-language formulations, we believe that (II), which states the absolute version of strong supervenience, is both much easier to understand,[17] and deeply independently intuitively compelling in its own right, in cases for which supervenience claims are often made, such as the supervenience of the moral on the non-moral.

What it says, in the moral case, is that for any moral property and any maximal non-moral property—that is, any complete way that something could be in every non-moral respect—if it is possible for something to have had both properties, then necessarily, anything which is that same way in every non-moral respect would also have to have that moral property. For example, if it is possible for an action to be wrong under some maximally determinate set of circumstances, then necessarily any action in the very same circumstances would be wrong. This claim is highly intuitive in its own right.

5.6.1 The Negations of Absolute Supervenience Theses are Independently Deeply Implausible

Moreover, we also believe that the intuitive compellingness of (II) for cases like the moral/non-moral case is even clearer, when we turn to look at its negation. Negating (II) and simplifying the resulting formulas by pushing the negation through (until all the negation-signs of the formula are contained in some atomic subformula) gives us:

[16] For example, see Paull and Sider [1992, 834] and McLaughlin [1997, 210].

[17] The difficulty in being sure that we understand (I) properly when we consider it in ordinary language, lies in being sure that the second 'necessarily' is clearly understood as having scope inside the first. Any reading on which this is not clear may simply be confusing world-relative strong supervenience with absolute strong supervenience.

Neg-(II) There is a B-maximal property which, for some A property F, could have been co-instantiated with F and could also have been instantiated without F.

We think that Neg-(II) is both particularly easy to understand and *very* implausible, for ordinary cases in which supervenience theses are maintained, such as that of the moral—in particular, of wrongness—on the non-moral. Observe:

Wrong: There is some complete characterization of an action in absolutely every natural respect, such that it is both possible for an action to satisfy that characterization and be wrong, and possible for an action to satisfy that characterization and not be wrong, without any other natural difference between the two situations.

This is extremely unintuitive. We think that such a scenario is very odd—it seems to undermine the core intuition in which our commitment to the supervenience of the normative on the natural is grounded.

The moral so far, is this: even though there is such an important logical difference between world-relative and absolute strong supervenience theses, the absolute version is extremely compelling in cases like this one. But of course, what we've observed in the earlier part of the paper is that the payoffs for Wedgwood's resistance to the S5 modal logic require him to deny the absolute version of strong supervenience for the moral on the natural. So he is committed to the unintuitive result illustrated by Wrong.

5.6.2 Two Possible Responses?

Presumably, one natural response for Wedgwood is to claim that our reasoning rests on a subtle confusion. Perhaps he might argue that when we are finding SS^{abs} intuitively compelling, that is because we are really confusing it with SS^{WR}, which is the real truth in the neighborhood. This confusion, if we were really making it, could explain our intuitive judgment. Call this the Confusion Hypothesis.[18]

We agree that this is a good strategy for Wedgwood to try, but we think that it is particularly implausible, and for two reasons. First, we believe that Neg-(II), which is equivalent to the negation of SS^{abs}, is one of the easiest-to-understand ways available of formulating any supervenience thesis or its negation. It is very implausible to think that when we consider such an easy-to-understand sentence, what is really happening, is that we are mixing it up with the negation of SS^{WR}, and not least because the negation of SS^{WR} is rather hard to understand, in its own right. Here is our best pass at how to express the negation of (I) in a natural way:

Neg-(I) There could have been something with an A property, but which would have possibly not had that A property, even if it had had the same B-maximal property.

It is extremely implausible that our grasp of such an easy-to-understand sentence as Neg-(II) is mediated by confusing it with the difficult-to-grasp content of Neg-(I), as

[18] Thanks to an anonymous referee for reminding us how the assumption of (E) may be thought to mediate such confusion.

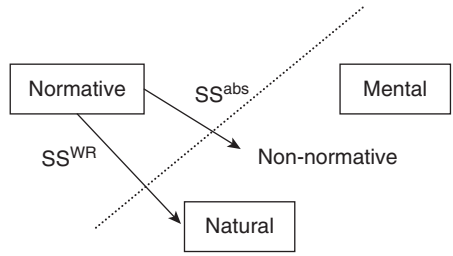


Figure 5.12

the Confusion Hypothesis would require. Moreover, our own plausibility-judgments persist when we stipulate that (E) fails—in fact, we don't even have to stipulate that (E) fails because we both do reject (E) independently of the intuitions discussed in this section and so the fact that our judgement is robust with respect to the rejection of (E) is evidence that the Confusion Hypothesis is false.

Alternatively, instead of appealing to the Confusion Hypothesis, Wedgwood could fall back on the important distinction between the thesis that the moral supervenes on the natural and the thesis that the normative supervenes on the *non-normative*. As he has reminded us in conversation, Wedgwood is not committed to denying the absolute strong supervenience of the normative on the *non-normative*—because he is inclined to hold that the class of non-normative properties includes mental properties. His picture therefore looks like figure 5.12.

As illustrated in figure 5.12, Wedgwood accepts that the set of normative properties absolutely strongly supervenes on the set of non-normative properties—provided that this set is understood to include intentional mental properties as well as strictly natural properties. He holds an analogous relationship to hold for the supervenience of the mental properties (figure 5.13).

Despite the acceptance of these absolute strong supervenience theses, however, this picture is still compatible with Wedgwood's non-reductive ambitions, because he is a non-reductivist about both the normative and the mental (figure 5.14).

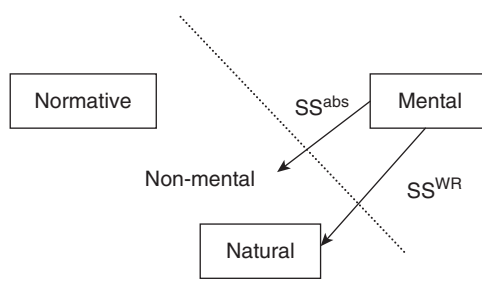


Figure 5.13

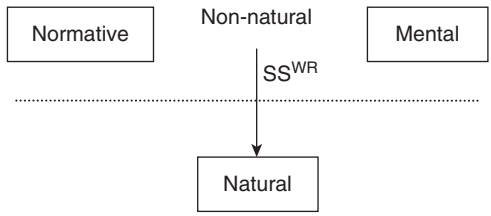


Figure 5.14

So instead of hypothesizing that we are confusing SS^{abs} with SS^{WR}, Wedgwood might claim that we are confusing the absolute strong supervenience of the normative on the non-normative (which is both plausible and true) with the absolute strong supervenience of the normative on the natural (which he claims is false, and only gains plausibility by confusion with the former).

It would be interesting enough to find that this sophisticated maneuver is required in order make good on taking advantage of relaxing S5 in order to accommodate supervenience without reduction. If this is right, then Wedgwood's views really will stand or fall as quite a large package, and it will be difficult to get the benefits of his idea of relaxing S5 without going whole hog on many of his other controversial ideas. But even so, though this response complicates matters, we find it ultimately unpersuasive. It is true that as we have formulated **Wrong**, above, it is formulated in terms of the supervenience of wrongness on the non-normative. But we could just as well formulate it in terms of the natural, which we continue to find just as unintuitive:

Wrong* There is some complete characterization of an action in absolutely every natural respect, such that it is both possible for an action to satisfy that characterization and be wrong, and possible for an action to satisfy that characterization and not be wrong, without any other natural difference between the two situations.

Against this, one might claim that **Wrong*** is only unintuitive if property-dualism is false. To see the force of this objection suppose that phenomenal properties are non-natural properties. Now, consider a scenario in which Mary viciously hits John. If whether this action is wrong may depend on how much pain it causes, and if pain is a phenomenal property, not a natural property, then **Wrong*** could be true.[19]

Two comments are in order: First, Wedgwood seems to be committed to denying absolute strong supervenience even if phenomenal properties are included in the supervenience base because not only does he think that the normative cannot be reduced to the natural, he also thinks that the normative cannot be reduced to the natural cum phenomenal. The normative is not reducible to any set of properties unless that set includes intentional properties and so—given what we have argued in

[19] Thanks to Tristram McPherson for pushing us on the relative implausibility of **Wrong** and **Wrong*** along these lines.

section 5.5.2—it seems to us that Wedgwood is committed to SS^{abs} being false for any supervenience-base that does not include intentional properties.[20]

Second, even though we wish to remain somewhat cautious about this point, we are both very sympathetic to a kind of naturalism which entails that phenomenal properties are natural properties, so we happen to think that the kind of property-dualism that the objection presupposes is most probably false (in fact, necessarily false). Consequently, we think it is anything but costless to show that the uninituitiveness of **Wrong*** can be explained away or resisted.

In conclusion, despite the allure of the idea that relaxed assumptions about modal logic can resuscitate non-reductive theories and keep at bay a variety of important arguments based on supervenience, the moral is: not so fast. All of these potential benefits also turn on the idea that absolute versions of supervenience theses are false and only their world-relative versions are true. But if what we've argued here is on the right track, the absolute versions of plausible supervenience theses are intuitively compelling in their own right.[21]

Appendix A.1 Proof that ii. and SS^{abs} are equivalent

We begin by noting that SS^{abs} can be broken up into the following two parts (i.e., that it is equivalent to their conjunction):

SS^{abs}_{1} $\quad \forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall x_{\in D(w)} \forall y_{\in D(v)} [\forall G_{\in B}(Gx \equiv Gy) \supset \forall F_{\in A}(Fx \supset Fy)]$

SS^{abs}_{2} $\quad \forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall x_{\in D(w)} \forall y_{\in D(v)} [\forall G_{\in B}(Gx \equiv Gy) \supset \forall F_{\in A}(Fy \supset Fx)]$

Observe that SS^{abs}_{1} and SS^{abs}_{2} are notational variants of one another (swapping x for y and w for v, and switching the order of the universal quanfiers and the biconditional). Hence, since SS^{abs} is equivalent to their conjunction, it is equivalent to each. So this reduces our problem to that of showing that ii. is equivalent to SS^{abs}_{1}.

ii. $\forall F_{\in A} \forall G_{\in B^*} [\Diamond \exists x(Fx \& Gx) \supset \Box \forall y(Gy \supset Fy)]$
$\Leftrightarrow$ (by possible-world equivalences for '$\Box$' and '$\Diamond$')

1 $\forall F_{\in A} \forall G_{\in B^*} [\exists w_{:R(@,w)} \exists x_{\in D(w)} (Fx \& Gx) \supset \forall v_{:R(@,v)} \forall y_{\in D(v)} (Gy \supset Fy)]$
$\Leftrightarrow$ (by quantifier movement rule from predicate logic)

2 $\forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall F_{\in A} \forall G_{\in B^*} [\exists x_{\in D(w)} (Fx \& Gx) \supset \forall y_{\in D(v)} (Gy \supset Fy)]$
$\Leftrightarrow$ (by second application of quantifier movement rule from predicate logic)

3 $\forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall x_{\in D(w)} \forall y_{\in D(v)} \forall F_{\in A} \forall G_{\in B^*} [(Fx \& Gx) \supset (Gy \supset Fy)]$
$\Leftrightarrow$ (by exportation and importation)

4 $\forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall x_{\in D(w)} \forall y_{\in D(v)} \forall F_{\in A} \forall G_{\in B^*} [(Gx \& Gy) \supset (Fx \supset Fy)]$
$\Leftrightarrow$ (reversing quantifier movement rule from predicate logic)

5 $\forall w_{:R(@,w)} \forall v_{:R(@,v)} \forall x_{\in D(w)} \forall y_{\in D(v)} [\exists G_{\in B^*} (Gx \& Gy) \supset \forall F_{\in A} (Fx \supset Fy)]$

[20] For this reasoning to go through we have to assume that intentional properties are not reducible to or constructible out of phenomenal properties.

[21] Special thanks to Billy Dunaway, Karen Bennett, and Ralph Wedgwood.

⇔ (by definition of B* (the set of B-maximal properties))

6 $\forall w_{:R(@,w)}\forall v_{:R(@,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$ QED

It is a trivial corollary of the equivalence of ii. and SS^{abs} that for any B model M, $M\vDash SS^{abs}$ just in case $M\vDash$ii.

Appendix A.2 Proof that for all B models M, $M\vDash SS^{WR}$ just in case $M\vDash$i

As with SS^{abs}, we begin by noting that SS^{WR} can be broken up into the following two parts (i.e., that it is equivalent to their conjunction):

SS^{WR}_{1} $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$
SS^{WR}_{2} $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fy\supset Fx)]$

Unlike SS^{abs}, however, SS^{WR}_{1} and SS^{WR}_{2} are not equivalent. (This is why we show only that in B models, $M\vDash$i just in case $M\vDash SS^{WR}$.) So we'll proceed by first showing that for any world @, SS^{WR}_{1} holds at @ just in case i. does. So if $M\vDash SS^{WR}$, then $M\vDash$i, giving us one direction, and if $M\vDash$i, then $M\vDash SS^{WR}_{1}$, giving us one half of the other.

SS^{WR}_{1} $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$
⇔ (by the definition of B* (the set of B-maximal properties))

1 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}[\exists G_{\in B^*}(Gx\&Gy)\supset\forall F_{\in A}(Fx\supset Fy)]$
⇔ (by quantifier movement rule from predicate logic)

2 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[(Gx\&Gy)\supset(Fx\supset Fy)]$
⇔ (by exportation and importation)

3 $\forall w_{:R(@,w)}\forall v_{:R(w,v)}\forall x_{\in D(w)}\forall y_{\in D(v)}\forall F_{\in A}\forall G_{\in B^*}[Fx\supset(Gx\supset(Gy\supset Fy))]$
⇔ (reversing quantifier movement rule from predicate logic)

4 $\forall w_{:R(@,w)}\forall x_{\in D(w)}\forall F_{\in A}[Fx\supset\forall G_{\in B^*}\forall v_{:R(w,v)}\forall y_{\in D(v)}(Gx\supset(Gy\supset Fy))]$
⇔ (second application of quantifier movement rule from predicate logic)

5 $\forall w_{:R(@,w)}\forall x_{\in D(w)}\forall F_{\in A}[Fx\supset\forall G_{\in B^*}(Gx\supset\forall v_{:R(w,v)}\forall y_{\in D(v)}(Gy\supset Fy))]$
⇔ (by possible-world equivalence for '$\Box$')

6 $\Box\forall x\forall F_{\in A}[Fx\supset\forall G_{\in B^*}(Gx\supset\Box\forall y(Gy\supset Fy))]$
⇔ (by assumption that $\forall x\exists!G_{\in B^*}(Gx)$)

i $\Box\forall x\forall F_{\in A}[Fx\supset\exists G_{\in B^*}(Gx\&\Box\forall y(Gy\supset Fy))]$

What remains, then, is to show that if $M\vDash$i, then $M\vDash SS^{WR}_{2}$. We'll show this by showing that if $M\vDash SS^{WR}_{1}$, then $M\vDash SS^{WR}_{2}$.

Let @ be any world in M, w be any world such that R(@,w), and v be any world such that R(w,v). To establish $M\vDash SS^{WR}_{2}$, we seek to show that

$$\forall x_{\in D(w)}\forall y_{\in D(v)}[\forall G_{\in B}(Gx\equiv Gy)\supset\forall F_{\in A}(Fy\supset Fx)].$$

Since $M\vDash SS^{WR}_{1}$, we know that SS^{WR}_{1} holds at v—i.e., that (using a and b as bound variables over worlds)

$$\forall a_{:R(v,a)} \forall b_{:R(w,b)} \forall x_{\in D(a)} \forall y_{\in D(b)} [\forall G_{\in B} (Gx \equiv Gy) \supset \forall F_{\in A} (Fx \supset Fy)].$$

Moreover, since we are assuming that M is a B model, we know that R(v,v) and R(v,w). Hence, (substituting v for a and w for b and applying universal instantiation):

$$\forall x_{\in D(v)} \forall y_{\in D(w)} [\forall G_{\in B} (Gx \equiv Gy) \supset \forall F_{\in A} (Fx \supset Fy)].$$

But this is a notational variant on what we want to show: switching x and y:

$$\forall y_{\in D(v)} \forall x_{\in D(w)} [\forall G_{\in B} (Gy \equiv Gx) \supset \forall F_{\in A} (Fy \supset Fx)] \quad \text{QED.}$$

6

The Price of Supervenience

My goal in this paper is to bring two things together. The first is an important contemporary modal challenge to non-reductive moral realism which I call the *explanatory argument from supervenience*. And the second is a package of ideas from Richard Price's *A Review of the Principal Questions in Morals* about the nature of moral facts. Although these two topics may seem unlikely bedfellows, Price's *Review* is one of the most full-throated and important historical defenses of non-reductive moral realism. Like many contemporary defenses of non-reductive moral realism, it emphasizes the analogy between the moral and the mathematical.[1] But *unlike* contemporary authors, the primary focus of Price's emphasis on the analogy between the moral and the mathematical is not their ontological commitments, or their epistemology. What Price sees as the primary analogy between the moral and the mathematical is their *modal status*: that both are domains of necessary truths.[2] This makes Price a natural interlocutor for contemporary *modal* arguments against non-reductive moral realism—and that is the role he will play in what follows.

So here is the plan: I'll start, in section 6.1, by introducing the modal challenge to non-reductive realism on which I aim to focus in this paper. Then in section 6.2, I'll explain why the analogy with mathematics might—at least in principle—be helpful to the non-reductivist in defending against at least one important form of this challenge. In sections 6.3 and 6.4, I'll show that this analogy opens up an interesting strategy for the non-reductivist—but that carrying out this strategy is far from trivial. This is where Price comes in. In section 6.5, I'll explain why Price held that all moral truths are

[1] Compare Putnam [2004], McGrath [2010], Enoch [2011], Parfit [2011], and Clarke-Doane [forthcoming], as well as further references in Clarke-Doane [forthcoming].

[2] "Right and wrong, it appears, denote what actions *are*. Now whatever any thing *is*, that it is, not by will, or decree, or power, but by *nature and necessity*. Whatever a triangle is, that it is unchangeably and eternally. [. . .] The natures of things then being immutable; whatever we suppose the natures of actions to be, they must be immutable. If they are indifferent, this indifference is itself immutable, and there neither is nor can be any one things that, *in reality*, we *ought* to do rather than another. The same is to be said of right and wrong, of moral good and evil, as far as they express *real characters* of actions. They must immutably and necessarily belong to those actions of which they are *truly* affirmed." [Price 1994, 50]. Though this passage illustrates the idea from Price that will be important for me in what follows, however, it oversimplifies Price's perspective, for Price also expresses related concerns about causal modality. I'll ignore these complications

necessary, and what commitments are involved in maintaining this claim. And finally, in sections 6.6 and 6.7 I'll put all of the pieces together and explore both the prospects and limitations of the resulting Pricean strategy of responding to the modal challenge.

Lest there be no misunderstanding, let me be clear up front that my aim is not to defend an answer to the modal challenge on behalf of the non-reductive moral realist. I do not actually think that non-reductive moral realism is true, and one of the most pressing worries for me, at least personally, about non-reductive moral realism is precisely the kind of modal challenge that I will be considering in this paper. The paper is, rather, an exercise in doing as well as I can on behalf of a view with which I actually disagree—using the tools I find to be the most promising, and aiming to achieve the standards that would satisfy a critic like me. To foreshadow, my conclusion will be that the Pricean strategy is well worth serious attention, offering striking but limited prospects for progress on this front. Its limitations help to shed light on the force and scope of the modal challenges facing non-reductive moral realism.

6.1 The Explanatory Argument from Supervenience

The modal challenge to non-reductive moral realism with which I will be interested in this paper is what I call the *explanatory argument from supervenience*. The argument starts, as its name suggests, with the observation that moral facts supervene on the non-moral.[3] There are many different ways of formulating supervenience theses more precisely, but at bottom, supervenience is the idea that there can be no difference in the moral facts without some difference in the non-moral facts.[4] However we precisify this idea, what is of interest to the explanatory argument from supervenience is that it requires necessary connections between the moral and the non-moral.

It is important, at the outset, to distinguish the explanatory challenge that I will be concerned with from the more familiar *direct* argument from supervenience to reduction. Both Jaegwon Kim [1984] and Frank Jackson [1998], and following them a number of other authors, have contended that supervenience theses of the sort that are plausible in metaethics *entail* reduction.[5] The arguments that this is so consist primarily

in what follows, but interested readers should consult Price's remarks about Newton in chapter 1 of the *Review*.

[3] For a prominent complaint in the broad family of the explanatory argument from supervenience, see Blackburn [1973], [1984], and [1985]. Blackburn believes that he has an argument against both reductive and non-reductive moral realism, however, and his argument differs from the ones considered here in several important ways. For the point, in a different context, that supervenience requires explanation, see especially Schiffer [1987, 153–154]. For versions of the explanatory argument from supervenience, see Schroeder [2007a, chapter 4] and Chapter 5, this volume. Scanlon [2009] considers a version of this argument and offers a response; we'll consider his response in section 6.4.

[4] See especially the papers in Kim [1993], as well as Bennett [2004].

[5] For versions of the direct argument, see Kim [1984], Jackson [1998], Streumer [2008], [2011], and Brown [2011].

in a technical component, paired with a strong thesis about the coarse-grainedness of properties, according to which properties can be distinguished only up to their intensions, and so properties that are necessarily shared are identical.[6] This assumption about the coarse-grainedness of properties is strong and in my own view, not particularly plausible. So my own view is that these arguments do not present a forceful challenge to non-reductive moral realism. Regardless of how forceful we take this problem to be, however, the explanatory argument from supervenience is a challenge that still remains, even once we respond to these direct arguments.

In contrast to the direct arguments, the explanatory argument from supervenience doesn't even need the assumption of full-fledged supervenience. It only needs one necessary connection between the moral and the non-moral to get off the ground. And indisputable such necessities are not hard to come by. To construct one, take your most cherished moral belief, and consider whether it could have been false even while every non-moral fact stayed the same. For me, I'm as sure as I am of just about anything that the fact that Carol Schroeder is my mother is not a reason for me to torture her. So the idea that this could have been false even while every non-moral fact was as things actually are strikes me as about as absurd as any. So for me, at least one necessity connecting the moral to the non-moral is intellectually rock-bottom, and I'm confident that some such necessity is rock-bottom for you, too. This necessity is a special case of supervenience: it is the thesis that some moral difference *couldn't* happen without any non-moral difference.

Once we get this far, the explanatory argument from supervenience can be pressed in two different ways. The first way of pressing the argument relies on a principle sometimes known as *Hume's Dictum*, which says that there are no necessary connections between distinct existences.[7] If supervenience means that there are necessary connections between the moral and the non-moral, then Hume's Dictum and supervenience together imply that the moral and the non-moral are not wholly distinct. But non-reductive realism is, on the face of it, the thesis that the moral and the non-moral are distinct realms of truths. So supervenience together with Hume's Dictum appear to create trouble for non-reductive realism.

The 'Hume's Dictum' way of pressing the explanatory argument from supervenience is tidy, because it appears to give us a direct argument against non-reductive moral realism, at least one of whose premises is bound to be indubitable. We'll see, over the course of this paper, that this tidy appearance is somewhat misleading. But I call attention to its tidiness, because this contrasts with a second way of pressing the explanatory argument, which is much less direct. Rather than relying on Hume's Dictum, which is both a completely general principle and only applies to some necessities—those that

[6] For a different take on the commitments of the direct argument from supervenience to reduction, see chapter four of Dunaway [2013].

[7] For more on Hume's Dictum, see Wilson [2010]. I'll worry later about how to formulate Hume's Dictum a little bit more precisely.

involve a *connection* between the moral and the non-moral—the second way of pressing the explanatory argument relies instead on the principle that necessities require explanation. Since supervenience commits us to necessities, it follows from this principle that these require explanation. But it is not so easy to explain these necessities in non-reductivist-friendly terms, the proponent of this argument alleges, without appealing to further necessities that are not explained. So the second form of the argument is structured more as an alleged-to-be-insurmountable challenge, rather than as a direct argument.[8]

Although Hume's Dictum doesn't itself mention explanation, it is natural to think of it as motivated as a special case of the idea that necessities require explanation—together with the idea that all necessities are explained by the lack of full distinctness among the entities involved. Take, for example, the necessity that something cannot be a red square without being red. As a necessity, this is a strong claim, because it rules out even the *possibility* of a non-red red square. But it is not puzzling why this is necessary—it is necessary because being red is just part of what it is to be a red square. This example both illustrates how a necessity can be explained, and why at least this sort of explanation seems to depend on a lack of distinctness between the entities involved, and correspondingly why it seems to be the sort of explanation that a non-reductive realist could not give of the necessities involved in supervenience.

Similar points go for other famous necessities, such as the impossibility of water that is not composed of H_2O. The necessity of there being no water that is not composed of H_2O is not brute; it strikes us immediately as something that stands in need of explanation. And the identity of water with H_2O is just the sort of thing to explain it. Since being water *is* being composed of H_2O, there is no puzzle about why it is impossible for there to be water that is not composed of H_2O. But this explanation again adverts to the lack of distinctness of the entities involved, and again it is clear that it is not the sort of thing that moral non-reductivists could appeal to.

In these remarks, I've been cherry-picking examples in order to provide intuitive support for the high-level principle that necessities require explanation. But of course this principle is not indisputable; indeed, some examples are much harder. Among the more interesting cases are the necessary connections involved in the determinate/determinable relationship. It is impossible for something to be crimson without being red, or red without being colored, but it's unlikely that being red is just being colored plus something else, in the way that being a red square is being red plus something else. Nor does it seem likely that redness is part of what it is to be colored. Now to evaluate whether the determinate/determinable relationship is a counterexample to Hume's Dictum, we would need a more fine-grained way of assessing whether determinables and their determinates count as 'distinct existences'. But at least this much is true: the distinctness of the moral from the non-moral posited by non-reductive realists—the

[8] See particularly chapter 4 of Schroeder [2007a] for a statement and defense of this form of the challenge.

sort of distinctness which is motivated by the thought that moral properties are simply "much too different" from non-moral properties to be analyzed in terms of them—does not seem to be consistent with the idea that moral properties are simply determinables with non-moral determinates.[9] So however these sorts of necessities work, they don't seem to be a promising model for non-reductive realists to explain their necessities, either, at least if their non-reductive realism is motivated, in part, by the thought that the moral is simply 'too different' from the non-moral.

So there seems to be at least some merit to at least the intuitive thoughts that necessities require explanation, and that failures of complete distinctness are at least a particularly promising, if not the unique, path to explaining such necessities. However, it is not my interest, here, to defend the force of this form of the explanatory argument from supervenience. For now, my aim is only to clearly distinguish these two forms the explanatory argument may take. Later we'll return to reconsider which of these forms poses the greater threat to non-reductive moral realism.

It is important to note that there is at least one very serious response in the literature to the explanatory argument from supervenience. In the course of developing a response to the direct arguments from supervenience to reduction, Ralph Wedgwood [1999], [2007] develops an ingenious set of commitments which taken collectively offer a way of responding to both forms of the explanatory argument from supervenience, without simply dismissing the need for explanations for necessities. By rejecting the modal principles known as S4 and S5, Wedgwood makes it natural to reject Hume's Dictum but replace it with a close and even more compelling neighbor principle, as well as to explain necessities by adverting to contingencies. Schmitt and Schroeder [2011] explore the intricacies of Wedgwood's ingenious view in detail and show how many moving parts it needs in order to be successful; my project in this paper is motivated by interest in how well the non-reductive realist can respond to the explanatory challenge without taking on the whole package of Wedgwood's commitments.[10]

6.2 Why the Analogy with Mathematics Might be Helpful

So much for the explanatory argument from supervenience. Our next task is to see why the analogy with mathematics might be a helpful one. The best way to see this is to start with the observation that the mathematical supervenes on the non-mathematical. All

[9] For the 'much too different' intuition motivating non-reductive moral realism, see especially Nagel [1986, 138], Huemer [2005], and Parfit [2011], as well as the discussion and further references in Schroeder [2005b].

[10] I'm interested, in particular, in what strategy the non-reductive realist might take to offer a response that consists in more than a gesture toward 'partners in crime'. As we'll see, the Pricean strategy in which I will be interested is inspired by the case of mathematics as a partner in crime, but the strategy goes beyond this point, by using the case of mathematics to construct a positive diagnosis of where at least one form of the explanatory argument may be evaded.

that supervenience requires, recall, is that there cannot be any difference of one kind without a difference of some other kind. And it is certainly impossible for there to be any difference in the truths of mathematics without there being a difference in something else. This is trivially true, in fact, because the truths of mathematics are necessities. Since it is impossible for necessities to be otherwise, it follows trivially that it is impossible for them to be otherwise unless something else was different as well.

Of course no one concludes from this case that the mathematical must reduce to the non-mathematical—and with good reason. Since math supervenes equally well on music and on biology, such an argument would have an equal claim to establish that math reduces to music as that math reduces to biology. (Here I assume that neither music nor biology reduces to the other, and hence that math cannot reduce to both, so I infer that this is a *reductio*.) So since the explanatory argument in metaethics also starts with supervenience and aims to get us to reduction, it would behoove us to think carefully about the case of mathematics.

The answer is given by the fact that the supervenience of the mathematical on anything else is *trivial*, in that it is guaranteed by the necessity of mathematics alone. It follows directly from this fact that though the supervenience of the mathematical involves some necessities, it does not involve any necessary *connections* between the mathematical and anything else. A different way of putting this is by saying that in cases of trivial supervenience, supervenience does not involve *co-variation*. The mathematical truths do not vary in company with the musical or biological truths in any way; rather, they supervene precisely because they do not vary at all.

On the face of it, in contrast, moral supervenience is not like this. Many important moral truths are not necessary, but contingent. For example, it would be wrong for me not to show up for the tenth annual Wisconsin Metaethics Workshop. But this is not necessarily true—it is true only because I promised to attend, and hence could easily have been false, if only I had been more selective about the commitments that I took on. The importance of the contingency of many of the most important moral questions about what we ought to do is a central theme in Sidgwick's *The Methods of Ethics*.[11]

The fact that many important moral truths appear to be contingent, rather than necessary, means that when the moral supervenes on the non-moral, it does so in a way that involves genuine co-variation, and not just the triviality of the necessity of the moral truths. And that makes it look like the example of the mathematical is not going to be very much help to the non-reductive realist. But the reason that Price will be of interest to us, is precisely that this is what he appears to deny. Price seems to think that moral truths are all *necessary*:

> Thus, then, is morality fixed on an immoveable *basis*, and appears not to be, in any sense, *factitious*; or the *arbitrary production* of any power human or divine; but *equally everlasting and necessary* with all *truth* and *reason*. [Price 1994, 52]

[11] In the *Methods*, this claim is closely related to the distinctive claim of intuitionism to offer a method for moral inquiry that is substantially non-empirical, in contrast to egoism and universalistic hedonism.

An analogy may help us to understand how Price seems to be thinking about things. To get started, consider *conjunctions* of mathematical and non-mathematical claims, such as that 2 + 2 = 4 and there are at least five planets in the solar system. Conjunctions like this one are *partly* mathematical, but partly not. Let's call them *bastard* mathematical claims, and contrast them to the *pure* mathematical claims. Conjunctions like this one will co-vary in truth along with the non-mathematical claim that there are at least five planets in the solar system, and hence they will supervene on the non-mathematical. But there is no puzzle about how they do so—this follows simply from the fact that the truths of *pure* mathematics are necessary. Since the truths of pure mathematics are necessary, the truth of their conjunctions with non-mathematical claims depends only on the truth of the non-mathematical conjuncts.

Conjunctions of pure mathematical claims with non-mathematical claims are not particularly interesting, but note that conjunctions of claims from each of two domains are just a special case of *relations* between the domains. In general, we may include in the domain of bastard mathematics everything that posits a non-trivial relationship between the pure mathematical facts and the purely non-mathematical facts. For example, the fact that the number of planets closer to the sun than earth plus the number of planets whose orbits are between those of Venus and Jupiter is equal to the number of gas giants is a truth of bastard mathematics. It's not a conjunction between any particular mathematical claim and any particular non-mathematical claims, but it requires the two to be lined up in the right way. And its truth depends on the truth of a particular (necessary) truth of pure mathematics—namely, that 2 + 2 = 4.

We may make two important observations about bastard mathematics. First, the truths of bastard mathematics supervene on the non-mathematical, and not simply trivially, in virtue of being necessities themselves. For some of the truths of bastard mathematics are contingent.[12] But second, there is no puzzle about how this could be. The supervenience of the bastard mathematical on the non-mathematical follows immediately from the fact that the truths of pure mathematics are necessary. For given that the truths of pure mathematics are necessary, there is only one way for the truth of bastard mathematical claims to vary: by the truth of their non-mathematical component varying. It is these two facts that our Pricean strategy for responding to the explanatory argument from supervenience will exploit.

6.3 A Path to a Solution?

So the trick that we want to pull off, on behalf of the non-reductive realist, is to respond to the explanatory argument from supervenience by exploiting this analogy. The elements of success are straightforward: we must be able to distinguish, among

[12] Note: not all. Bastard mathematical claims that relate the mathematical to non-mathematical non-contingencies will be necessarily true or necessarily false.

the moral truths, between those which are *purely* moral, and those which are merely 'bastard' relationships between the pure moral facts and the non-moral facts. In addition to being intuitively *purely* moral, the pure moral truths must all be necessities. Their supervenience on the non-moral will therefore be trivial, on analogy with the supervenience of the truths of pure mathematics. The bastard moral truths will also supervene on the non-moral, but this will be no more puzzling than how the bastard mathematical truths supervene on the non-mathematical.

Such a distinction between 'pure' and 'bastard' moral claims allows us to make sense of Price's claim that all moral truths are necessary—for this is true of the genuine, *pure* moral claims, and though it is not true of the bastard moral claims, even Price will allow that relationships between the moral and the non-moral may be contingent. Consequently, I will call this divide-and-conquer approach to the explanatory argument the *Pricean strategy*. There is much more to be said about exactly what the distinction between pure and bastard moral claims consists in, so the Pricean strategy can be implemented in different ways, depending on how we answer that question. In what follows, we'll consider three such implementations. But the Pricean strategy in general is just the basic idea inspired by the analogy to the case of mathematics.

In the mathematical case, our examples of 'bastard' mathematical claims were gerrymandered claims that we were not intuitively inclined to classify as mathematical, to begin with. If there is a similar division among moral claims between 'pure' and 'bastard', however, it will be less obvious. Claims such as that it is wrong for me to fail to show up for the tenth annual Wisconsin Metaethics Workshop are intuitively paradigmatic moral claims. Yet if we are to exploit our observations about the mathematical case in order to respond to the explanatory argument from supervenience on behalf of the non-reductive realist, it is claims like these that are going to have to be classified as the analogue of the 'bastard' mathematical claims. So if there is a division among moral claims between the 'pure' and the 'bastard', it will be substantially less obvious than in the case of mathematics.

Though this distinction is going to be less obvious, moreover, it is very important exactly how we draw the line between these two classes of intuitively moral claims. For in classifying ordinary moral claims such as that it would be wrong of me to fail to show up for the tenth annual Wisconsin Metaethics Workshop as merely 'bastard' claims, the envisioned response to the explanatory argument from supervenience is effectively going to give up on the idea that these claims are wholly distinct from the non-moral. They are still not *completely* reducible to the non-moral, but that is simply in virtue of the fact that they relate the non-moral to a special class of 'pure' moral claims which are, in fact, wholly distinct from the non-moral. So in order to adopt this strategy, the non-reductivist must restrict her claims about the distinctness of the moral and the non-moral to the privileged, 'pure' class of moral claims. It must be these claims, and their special status, that make all moral claims intuitively moral.

I want to emphasize how important this is. The non-reductivist starts with the intuitive view that moral claims are wholly distinct from non-moral claims. But in order

to exploit the analogy with mathematics, she must restrict this thesis—only *some* moral claims, it turns out, are wholly distinct from non-moral claims, while others are not. If this move is to preserve the spirit of the non-reductivist's original view, it had better turn out that the class of 'pure' moral claims has some claim to be what the non-reductivist most cared about holding to be wholly distinct from the non-moral, in the first place. They must be the sort of claims which it makes sense to think are so central to morality, that 'bastard' moral claims seem to count intuitively as moral precisely because of their relationship to them. As we investigate various ways of trying to make good on this strategy in what follows, this is one of the most important constraints for us to keep in mind.

Before going forward, it is essential to understand two important things about the Pricean strategy. First, it constitutes a response to the first form of the explanatory argument, because it is wholly consistent with Hume's Dictum. The basic form of the response is to divide the moral truths into two categories: the 'pure' moral, which are wholly distinct from, and hence irreducible to, the non-moral, and the 'bastard' moral, which are not, after all, wholly distinct from the non-moral, because they are defined, in part, in terms of non-moral claims. Both of these domains supervene on the non-moral, but only one involves a necessary connection. The supervenience of the bastard moral on the non-moral requires a necessary connection, and so Hume's Dictum applies, but it is innocuous, because the non-reductivist grants that these are not, after all, distinct existences. On the other hand, the supervenience of the pure moral on the non-moral does not involve any necessary connections—it follows trivially from the necessity of the pure moral truths, as in the case of mathematics. And so Hume's Dictum does not apply.

This observation leads immediately to our second: that this line of response only gives us a response to the first form of the explanatory argument—the one which depends on Hume's Dictum. This is because both parts of the account turn on the assumption that pure moral claims are all necessary—this is both what explains the supervenience of the bastard moral on the non-moral, and what explains why the supervenience of the pure moral on the non-moral is trivial. But now this necessity stands in need of explanation, according to the second form of the explanatory argument. Still, having a response even to the first form of the argument constitutes some real progress, so let's see what we can do with it.

One first thought that you might have, at this point, is that once we see the structure of this Pricean strategy, it is in some way trivial to take advantage of it. This thought is a natural one, and one way of motivating it is by way of the idea that since any supervenience thesis is committed to necessary truths, we can simply take those necessities to be the 'pure' moral truths. But unfortunately, it is not quite so easy to do this as it sounds. Because it's important to see that implementing the Pricean strategy is a non-trivial task, I'm going to take the remainder of this section to walk through this problematic reasoning in some detail. We may think of it as one way of implementing the Pricean strategy—the *trivialization* implementation.

First, the reasoning that might lead us to think that the Pricean strategy is trivial. We may start, to make things precise, by assuming a particular kind of supervenience thesis, *strong* supervenience. According to strong supervenience, no two possible entities (whether or not they exist at the same world) differ in any moral property without also differing in some non-moral property. So for any possible action *x* that is wrong in world *w*, there is a necessarily true conditional, F*x* → (*x* is wrong), where 'F' is a complete non-moral characterization of *x* in *w*—the conjunction of all of its non-moral properties.[13] This (material) conditional is guaranteed to be necessarily true, because by strong supervenience, there is no possible entity in any possible world which shares all of *x*'s non-moral properties (i.e., which satisfies 'F') but differs in some moral property (for example, in whether it is wrong). So now construct such a conditional for every pair <*x*,*w*> such that *x* is wrong at *w*, and let B be the (infinite) conjunction of all of these conditionals. By construction, B is also necessary. Similarly, let A be the disjunction of the predicates 'F' in the antecedents of each of these conditionals. By construction, A is completely non-moral.

Finally, compare the claim 'A*x*&B' to the claim '*x* is wrong'. Since every disjunct 'F' of the predicate 'A' is the antecedent[14] of a necessary conditional whose consequent is '*x* is wrong', 'A*x*&B' trivially has '*x* is wrong' as a necessary consequence. But since we have a disjunct in 'A' for every possibly wrong action, and 'B' is necessary, '*x* is wrong' also has 'A*x*&B' as a necessary consequence. So our construction guarantees that '*x* is wrong' and 'A*x*&B' are necessarily equivalent. But since B is necessary (and moral) and 'A' is completely non-moral, this means that we've succeeded at reconstructing an ordinary moral claim like '*x* is wrong', which as we've already noted, seems moral but can sometimes express a merely contingent truth, as a conjunction of a necessary moral claim and a possibly contingent wholly non-moral claim. And that is the core of what our Price-inspired strategy requires us to do—it requires us to defend the view that ordinary, contingent moral claims are really just bastard moral claims, and conjunctions are the simplest case of bastard moral claims.

So far, so good. At this point we've seen enough to see why the thought that the Pricean strategy might somehow be trivial to carry out might be initially tempting. But I don't think that this can be right, for at least two different reasons. Both turn on the fact that the Pricean strategy that I've been outlining actually requires *more* that we've gotten so far. To carry it out, it is not enough to construct, for each contingent moral claim, a conjunction of contingent non-moral claims and necessary moral claims with which it is necessarily equivalent; two more things are required. First, this conjunction must also reveal the true 'bastard' nature of the contingent moral claims, by revealing to us what they are really about. And second, the necessary moral claims which appear

[13] The reader will oberve that the construction in this and the following paragraph has much in common with the techniques applied in Kim [1984] and Jackson [1998].

[14] Throughout the statement of this argument I'm being slightly sloppy; strictly speaking, for every disjunct of the predicate 'A', the corresponding sentence 'Fx' is the antecedent of a necessary conditional, etc.

in these conjunctions must be plausibly construed to be *pure* moral claims. But I don't think that this construction gets us either of these things.

To see why not, recall how we constructed 'B'. We chose it to be a conjunction of infinitely many material conditionals, each of whose antecedents is the conjunction of all of the non-moral properties of some action at a world at which that action is wrong, and each of whose consequents is '*x* is wrong'. It strikes me as deeply implausible that we've elucidated the true nature of 'wrong' claims by analyzing them in terms of . . . wrongness. So though we've constructed a claim that is necessarily equivalent to 'x is wrong', I don't think that we've shown that 'x is wrong', itself, is a bastard claim, even if we assume that 'B' is a pure moral claim.

It is true that this objection requires a fine-grained conception of properties—the idea that different properties can be necessarily shared by the very same things. And this is controversial—indeed, the direct arguments from supervenience to reduction work by denying it. But for precisely that reason, I think that it ought to be safe in this context. For if necessarily equivalent properties must be identical, then the non-reductive moral realist should be worried about the direct arguments from supervenience to reduction, and the explanatory argument is otiose.

However, the second problem with this construction doesn't require a fine-grained conception of properties. It is that though the construction gives us every reason to think that 'B' is a moral claim, and that it is necessary, it gives us no reason whatsoever to think that it can plausibly be construed as a *pure* moral claim. For by construction, it is simply the conjunction of infinitely many material conditionals, each of whose antecedents is wholly non-moral. Nothing about this makes it at all clear how 'B' could really be the 'pure' moral claim that non-reductive realists need in order to carry out our Price-inspired strategy.

Of course, it *could* be that this infinite conjunction of material conditionals is itself, despite appearances, a pure moral claim. Or if not, it is at least possible that it could be necessarily equivalent to a pure moral claim, which would be just as good for our purposes. Nothing that we have said can rule this possibility out. But on the other hand, nothing we have said has given us any insight, either, into why we should think that this is true. And this is why we need more from Price than the idea that pure or genuine moral claims are all necessary. For while there is a trivial path to seeing contingent moral claims as at least equivalent to conjunctions of necessary moral claims with contingent non-moral claims, there is no trivial path to seeing how those necessary claims could be *pure*. A deeper look at Price can give us a picture of exactly that. But first, it will help to explore a recent implementation of the Pricean strategy by T. M. Scanlon.

6.4 Being Realistic About Reasons

In the last section, we saw that though the Pricean strategy offers some *prima facie* promise to offer a response to the 'Hume's Dictum' form of the explanatory argument from supervenience, it is not trivial to carry this strategy out. To carry it out, we need

not only a diagnosis of how each contingent moral claim really consists in a relationship between a necessary moral claim and a contingent, wholly non-moral claim, but an account of what these necessary moral claims are which makes clear why they are the 'pure' moral claims about which we wanted to be non-reductivists, in the first place. Fortunately, in his 2009 Locke Lectures, *Being Realistic About Reasons*, Scanlon sketched a response to modal worries about non-reductive normative realism that exemplifies the Pricean strategy.[15] Scanlon's response has the virtue of advocating a particular conception of the pure normative necessities, and hence it is exactly the right sort of thing for us to look at, in order to see how we might try to improve on the trivialization implementation the Pricean strategy. Since it is the second implementation of the Pricean strategy that we will consider, I will refer to it as *Scanlon's* implementation.

As we've done here for morality, but generalizing to the case of the normative more generally, Scanlon distinguishes between 'pure' and 'mixed' normative claims, and like us, he takes disjunctions of pure normative claims with non-normative claims as a paradigm of 'mixed' claims. Then, discussing the case of supervenience, he reasons,

> To understand the phenomena of covariance and supervenience it is important to be clear what kind of normative claims they involve. The normative facts that can vary as non-normative facts vary are facts that consist in the truth of mixed normative claims . . .
>
> [Scanlon 2009, 28]

Here Scanlon is making the point that it matters which supervening claims *co-vary* with the non-normative in such a way as to involve a necessary *connection* of the kind that Hume's Dictum prohibits between distinct existences. According to Scanlon in this passage, only *mixed* claims co-vary in this way, which shows that in keeping with our Pricean strategy, Scanlon is assuming that the pure normative claims are all necessary.

In contrast to the trivialization implementation that we considered in the last section, Scanlon offers a positive conception of what these pure normative necessities are like—one which promises to give us some insight into why they are really pure normative claims:

> The essential element in normative statements is not a term referring to an entity, but a relation: the relation that holds between a proposition, a set of conditions, and an action or attitude when p is a reason for a person in situation c to do or hold a.
>
> This sounds like a three-place relation, but it contains an implicit universal quantification over agents: it holds that p is a reason for *any* agent in c to do a. And we want not only to make not only general claims of this kind but also to claim, of a particular agents x, that some fact p is a reason for him or her to do a in c. So the underlying relation must be a four-place one R(p, x, c, a). [Scanlon 2009, 19]

[15] So far, in keeping with Price's own discussion, I've been focused on the case of non-reductive *moral* realism. But Scanlon generalizes to the case of non-reductive *normative* realism more generally, and I will follow him in what follows.

We can see immediately from this description that Scanlon's proposal has at least some *prima facie* promise to help us in exactly the place where the trivialization implementation dropped the ball. It is a common idea about normativity, endorsed by many different philosophers—including myself on another occasion—that reasons are the fundamental normative concept, in virtue of their relationship to which all normative claims count as normative. So if it is widely accepted that reasons are essentially normative, and all other normative claims are normative by virtue of implicating reasons, then claims about reasons have a particularly strong claim to be 'pure', in the sense that we require. And that is what Scanlon is claiming—that the pure normative claims are claims about reasons. So his picture offers us precisely what the trivialization implementation does not—an answer to why the special category that he has picked out is genuinely pure.

Unfortunately, whereas the trivialization implementation could easily show that its special category of normative claims are *necessary*, this is something Scanlon needs to take more care about. For though it is plausible—or at least, widely agreed—that all normative claims are normative in virtue of implicating reasons, it is also clear that ordinary claims about reasons are often contingent, in the same way as ordinary claims about wrongness or about what someone ought to do. In fact, with reasons this is even easier to see, for in general, reasons are facts which support some course of action or other—but many facts are contingent. So, for example, one reason for you to keep reading this paper is that it is insightful and penetrating, but this can only be true because the paper is in fact incisive and penetrating, which, though true, could easily have been otherwise. If it were otherwise, then it wouldn't be true that one reason for you to keep reading this paper is that it is insightful and penetrating. So consequently, that ordinary claim about reasons is contingent. And the same goes for any ordinary reason attribution citing any contingent fact whatsoever as a reason, as Scanlon himself recognizes. Indeed, ordinary claims about reasons are one of his leading examples of *mixed* normative claims:

> The normative facts that can vary as non-normative facts vary are facts that consist in the truth of mixed normative claims, such as the claim that someone has a reason to do a certain action, or that a particular consideration is such a reason. [Scanlon 2009, 28]

Scanlon is not unaware of this problem; in fact, it is precisely because of this problem that he takes care to appeal to his four-place R(p,x,c,a) relation:

> The essentially normative content of **R**, however, is independent of whether p is true: it lies in the claim that, whether p is the case or not, if p *were* the case it would be a reason for someone in c to do a. [Scanlon 2009, 23]

So according to Scanlon, ordinary 'reason' claims in natural language are only 'mixed' normative claims—the only genuine pure normative claims are those of the form, R(p,x,c,a). And claims of *this* form are all necessary, even though ordinary 'reason' claims are not.

But now we have a new problem. For though it is plausible (or at least, widely agreed) that reasons are the central normative concept, it is not widely agreed that R(p,x,c,a) is the central normative relation. Indeed, I'm not entirely clear that I understand this relation at all. And Scanlon's explanatory gloss doesn't do a lot to reassure me that it is genuinely a pure normative relation. For he explains what this means by saying that its content comes from a subjunctive claim about whether p *would* be a reason if it were true.[16] But recall that ordinary 'reason' claims are *not* pure—they are *mixed*. And so now I think we have a problem: if we're making sense of the content of supposedly 'pure' normative claims by cashing them out in terms of 'mixed' normative claims, then I think we've lost the standing to claim that the 'pure' claims are the basic ones and the 'mixed' ones are just what we get when we mix pure normative claims with non-normative ones.

Perhaps this objection is unfair to Scanlon; he may have intended, in the passage I've just quoted, merely to be giving us some assistance in cottoning on to the relationship between ordinary 'reason' claims and the R(p,x,c,a) relation, rather than telling us *in what the essentially normative content of R lies*, as he explicitly says. But whether or not the objection is unfair to Scanlon, it's important to be clear that this strategy requires a defense not only of the division of the normative into 'pure' and 'bastard' or 'mixed' categories, but that this be done in a way that is plausible and defensible.

So far, we've seen that by appealing to reasons, Scanlon's account has some promise to be able to make good on the claim that his distinguished class of normative claims are genuinely 'pure'. But at the same time, we've also seen that there is a tension between this and the requirement that the pure normative claims all be necessary—for ordinary claims about reasons are not all necessary. But in fact, there is a deeper problem with Scanlon's answer to the modal challenge, which we may see by returning to reflect further on what, exactly, Hume's Dictum requires.

So far, I've been very sloppy about exactly how to formulate Hume's Dictum precisely. For our purposes, it's been sufficient to note that it forbids necessary co-variance between distinct realms of truths. But at a closer pass, we might think that it says something more general: such as that there is no relation R such that for distinct existences x and y, necessarily R(x,y). And if the correct version of Hume's Dictum says such a thing, then Scanlon's treatment clearly runs afoul of Hume's Dictum after all—for presumably p, x, c, and a are at least in general distinct existences, and Scanlon's basic normative necessities say that some relation—the R relation—holds of such tuples necessarily.[17] This reasoning suggests that the Pricean strategy comes with (you guessed it) a strict price: the basic normative properties must be *monadic*, rather than polyadic, if we aim to avoid violations of Hume's Dictum, so construed.

[16] Presumably he would say something similar to solve the closely related problem about what it means for p to be a reason for x to do a in c, in cases in which x is not actually in c. Since claims of the form R(p,x,c,a) are necessary, but whether a given agent x is in a given circumstance c is only contingent, the R(p,x,c,a) claims must often be true even though x is not in c.

[17] For this point I'm indebted to Krister Bykvist.

6.5 The Standard Model for Normative Explanations

So here is where we are so far: we know the general form of the Pricean strategy to respond to the Hume's Dictum form of the explanatory argument from supervenience. But we've also seen that it's non-trivial—indeed, *highly* non-trivial—to make good on all of the elements of this strategy. For to carry it out, we need some diagnosis of a privileged class of normative claims, such that all other intuitively normative claims can be understood as relationships between such claims and the non-normative. We need to have a firm enough grip on the content of these privileged claims that we are confident that they count as genuinely 'pure'. Yet this grip cannot come in a form that connects these claims to any of the ordinary classes of normative claims that are sometimes contingent, for they must all be necessary. And finally, these necessities themselves need to respect Hume's Dictum by not consisting themselves in necessary connections between distinct existences. And so in particular, they cannot be relational claims whose relata are distinct existences at all. If we are to take advantage of this Pricean strategy, therefore, we need to do better at meeting each of these criteria. Fortunately for us, a closer look at Price gives us just the tools that we need.

One of the most striking features of Price's *Review of the Principal Questions in Morals* is its intellectual indebtedness to Ralph Cudworth's *Treatise Concerning Eternal and Immutable Morality*. Despite citing Cudworth only on peripheral points and getting his title wrong,[18] Price returns over and over to a pattern of argument that to the best of my knowledge is first given in the first section of Cudworth's *Treatise*. I call it the *Cudworthy argument*, and have argued elsewhere that it is important enough to be well-worth chewing on again.[19]

The key background assumption that drives the Cudworthy argument is a general theory of how explanations of moral facts must work—a substantive picture that I call the *Standard Model Theory*. According to this picture, when we explain why something is wrong by adverting to some fact, this explanation always works by adverting to something else that is wrong, and pointing out that in virtue of our explanans, the action in question results in doing the other thing that is (as adverted) wrong.[20] In moral philosophy, we're familiar with such explanations all over the place. Having money is good because it lets you buy things. That's a good explanation because being able to buy things is good, and what the explanans tells us is that in virtue of having

[18] Price makes a point of always saying 'immutable and eternal', rather than 'eternal and immutable', including when he cites Cudworth's title [Price 1994, 55].

[19] See Chapter 1, this volume. The Cudworthy argument is also prominently featured as Clarke's main argument against Hobbes in his Boyle lectures (Clarke [1967]).

[20] In Chapter 1, this volume, I allowed that the historical Cudworth and Price each had views that were slightly more sophisticated than that I've described here, for reasons that are connected to the dialectical role of the early versions of the open question argument in each of their discussions. This complication won't matter for our purposes, here.

money, you are able to buy things. In this case, we say that what we explained is why having money is *instrumentally good*, and when we reach a point where such subsumptive explanations can go no further, we say that we have found something that is *intrinsically good*.

Cudworth and Price held that the same goes for explanations of what is wrong. It is wrong for me to fail to show up for the tenth annual Wisconsin Metaethics Workshop, because I promised to come. This is a good explanation, Cudworth claims, because what the explanans points out, is that in virtue of failing to show up for the workshop, I would break a promise, and it is wrong to break a promise. So failing to show up for the conference is only, as we might put it, instrumentally wrong, and we have only located something that is intrinsically wrong when such explanations can go no further.

Here is Cudworth presenting this picture:

> As for example, to keep faith and perform covenants is that which natural justice obligeth to absolutely. Therefore upon the supposition (*ex hypothesi*) that any one maketh a promise, which is a voluntary act of his own, to do particular something which he was not before obliged to by natural justice, upon the intervention of this voluntary act of his own, that indifferent thing promised **falling now under** something absolutely good and becoming the matter of promise and covenant, standeth for the present in a new relation to the rational nature of the promiser, and becometh for the time a thing which ought to be done by him, or which he is obliged to do. [Cudworth 1996, 20, boldface added for emphasis]

In this passage we see the main elements of the picture that I've just described. There is something that is absolutely obligatory—necessarily and immutably so. When we explain why something is obligatory by appeal to a further fact, what we are doing is explaining why, given that fact, this action *falls under* the case of what is absolutely obligatory.

I want to point out three very important features of this picture. First, it is highly intuitive to draw the line between 'pure' and 'bastard' moral claims between facts about intrinsic and instrumental value, and similarly, between intrinsic and instrumental wrongness. Second, it is a familiar assumption about intrinsic value (and similarly, about intrinsic wrongness, though this is much less discussed) that facts about such things are necessary, rather than contingent.[21] (I'm not endorsing this assumption; only pointing out that it is widely held.) And finally, the idea that intrinsic value is necessary is not a coincidence, *if* we accept Cudworth and Price's idea that all moral explanations must be subsumptive.

The reasoning here is not quite airtight, but it is simple and attractive: if intrinsic value were contingent, then it could turn out that whether something was good could depend on something else (otherwise we'd have a violation of supervenience). But if this something else could explain a difference in whether the thing is good, then by

[21] This assumption is most famously made by Moore [1903] and is the basis of his theory of organic unities. For an important dissenting opinion, see Korsgaard [1983]. Ironically (from my perspective), lecture one of Korsgaard [1996] advances a version of the Cudworthy argument, drawing on Clarke [1967].

the Standard Model Theory, it must do so by adverting to something else, more general, that is good, and that the thing in question is a way of getting in some cases, but not in others. So a subsumptive explanation needs to advert to a claim about what is good that encompasses all of the worlds which the explanation encompasses within its scope. Eventually, we must reach a violation of supervenience, or a case where there is a moral difference and a subvening non-moral difference but the non-moral difference does not do anything to explain the moral difference, or we must have an infinite sequence of explanations achieving greater and greater generality but never encompassing all possible worlds—or we must advert to something that is necessarily good. So if there are not to be unexplanatory difference-makers or infinite sequences of explanations, then things are going to have to end with something that is necessarily good. And since the claim about goodness that such explanations end with is the one about *intrinsic* goodness, it follows that claims about intrinsic goodness will be necessary. What this reasoning illustrates, is that Moore's idea that intrinsic value must supervene on intrinsic properties is not simply a linguistic error, confusing two senses of 'intrinsic'. Rather, it is the natural upshot of the Standard Model Theory as applied to 'good'.[22] Note again that I'm not saying that this conclusion is right—for I don't actually think that the Standard Model Theory is true. But I think the case is good that this is the right conclusion to draw if you accept the Standard Model Theory.

Where Price differs from Cudworth, is that he appears to think that bastard moral truths are not really moral truths at all. Even if you have promised to attend a workshop, failing to attend the workshop is not wrong, according to Price—what is true instead, is that *by* failing to show up for the workshop, you break your promise. And breaking your promise is wrong. So the apparent moral fact that it is wrong for you to fail to show up for the conference is not only explained by the fact that it is wrong to break your promises and the fact that you promised to attend the conference; it is actually an unholy hybrid of the fact that it is wrong to break promises and the fact that you promised to show up for the workshop.

Here is Price presenting his version of the picture:

> When an action, otherwise indifferent, becomes obligatory, by being made the subject of a *promise*; we are not to imagine, that our own will or breath alters the nature of things by making what is indifferent not so. But what was indifferent *before* the promise is still so; and it cannot be supposed, that, *after* the promise, it becomes obligatory, without a contradiction. All that the promise does, is, to alter the connexion of a particular effect; or to cause that to be an *instance* of right conduct which was not so before. There are no effects producible by us, which may not, in this manner, fall under different principles of morality; acquire connexions sometimes with happiness, and sometimes with misery; and thus stand in different relations to the eternal rules of duty. [Price 1994, 51–52]

So Price and Cudworth share the same underlying picture, but differ over how to classify what I've been calling bastard moral claims. Cudworth is content to allow that there

[22] Compare Schroeder [2009b] for further discussion.

are contingent moral obligations, so long as we recognize that for there to be any moral obligations at all, some must be necessary. So a non-reductivist who follows Cudworth would refine the statement of her non-reductivism, to apply only to *pure* moral truths, and not to moral truths in general (many of which are bastard). Whereas Price seems to want to insist that ordinary moral terms like 'obligatory' can only apply necessarily. A non-reductivist who follows Price would continue to maintain that all moral properties or truths are irreducible, but explain away the appearance that bastard moral claims are really moral claims. This, I think, is largely a terminological dispute. Either way, the core of the strategy drawing on the Standard Model Theory is simple. It is a direct defense of the move that we've seen has just the right sort of flavor to offer a response to the first version of the explanatory argument from supervenience to reduction.[23]

6.6 Putting it Together

Piecing together a natural version of Price's picture in order to develop a more properly Pricean implementation of the Pricean strategy, we get the following ideas: first, we may take *intrinsic wrongness* to be a property of action-types, as opposed to token acts. This is because token acts exist only contingently, but we want claims about intrinsic wrongness to be pure, and hence necessary. Second, we will assume that any action-type that has the property of *intrinsic wrongness* has it necessarily. However, third, 'wrong' is a binary predicate of action-types and agents. It expresses the relation that if the agent performs that action, then she will perform an action of a type that has the property, *intrinsically wrong*. So, for example, 'failing to show up for the tenth annual Wisconsin Metaethics Workshop is wrong for Mark' is contingent, even though intrinsic wrongness is necessary wherever instantiated. We could easily expand this picture to apply to other moral terms like 'good' and 'ought', but all of the lessons will mirror those from this case.[24]

Given this picture, the Pricean response to the first form of the explanatory argument from supervenience is exactly as we've anticipated: only atomic ascriptions of intrinsic wrongness and claims in their Boolean closure are *pure* moral claims. Ordinary claims about what is wrong, in contrast, are impure, bastard moral claims.

[23] Another way of thinking about the Pricean picture that I've been describing here is as a kind of arch-generalism. For discussion of this dimension of both Cudworth and Price, see Schroeder [2009b].

[24] Moore [1903] is surprisingly explicit about exactly this picture, in the passages of *Principia Ethica* between the Open Question argument and his introduction of organic unities: "Whenever we judge that a thing is 'good as a means,' we are making a judgment with regard to its causal relations: we judge *both* that it will have a particular kind of effect, *and* that that effect will be good in itself. . . There are, then, judgments which state that certain kinds of things have good effects; and such judgments, for the reasons just given, have the important characteristics (1) that they are unlikely to be true if, they state that the thing in question *always* has good effects, and (2) that, even if they only state that it *generally* has good effects, many of them will only be true of certain periods in the world's history. On the other hand there are judgments which state that certain kinds of things are themselves good; and these differ from the last in that, if true at all, they are all of them universally true. It is, therefore, extremely important to distinguish these two kinds of possible judgments. Both may be expressed in the same language . . ." [1903, sections 16–17].

Pure moral claims supervene trivially on the non-moral, in the same way as math, and without co-varying or being otherwise "connected" in any way. So there is no conflict with Hume's Dictum. In contrast, bastard moral claims do co-vary with claims about the non-moral, but by construction, they are not distinct existences. So again there is no conflict with Hume's Dictum.

This picture has promise to make good on each of our criteria: it *is* plausible that 'wrong' claims, in general, are about the relationship between what is intrinsically wrong and the non-moral facts. Claims about intrinsic wrongness, in turn, *are* a plausible candidate to be the genuinely 'pure' moral claims. And—at least conditional on the Standard Model Theory—it *is* at least plausible that claims about intrinsic wrongness are all necessary.

Our final requirement for a solution is that the 'pure' moral necessities must not themselves run afoul of Hume's Dictum, and over this issue we must take some greater care. In contrast to Scanlon's pure necessities, which consisted in the relation R holding necessarily among the distinct existences p, x, c, and a, the more strictly Pricean account to which we've appealed reserves the title of 'pure' for atomic attributions of intrinsic wrongness and their Boolean combinations—and intrinsic wrongness is taken to be a one-place property, rather than a relation. So at the least, we don't run afoul of this constraint in the same way that Scanlon does.[25]

Still, this may not put us in the clear. After all, intrinsic wrongness is itself presumably a distinct existence from any of the action-types which instantiate it. And so we might worry that if some action-type is intrinsically wrong, then that itself constitutes a necessary connection between distinct existences, in violation of Hume's Dictum. This may be right—if so, I think it would show the Pricean strategy to be ultimately fruitless. Still, I think there is reason to be optimistic on behalf of the non-reductive realist who seeks to employ this strategy. Fortunately for the prospective defender of the Pricean strategy, there are (regress-based) reasons to deny that property instantiation is itself a relation. And if it is not, then the mere fact that intrinsic wrongness is necessarily instantiated by the action-type of torturing puppies does not suffice for there to be some necessary relation between those distinct existences.

Could the Pricean strategy be committed to some other necessary relation, then? Here is one last attempt. The proponent of the Pricean strategy will note that it follows from the necessity of the intrinsic wrongness of torturing puppies that necessarily, if someone would torture puppies by moving their arm in a certain way, then if they

[25] Famously, this fits Moore's view of 'good' as expressing a monadic property. Can this strategy be extended to reasons? It can, if being intrinsically reason-supported is a one-place property of action-types, rather than a relation between considerations and the action-types they support. On such a view, the three-place 'reason' relation will be the bastard relation that holds of a fact p, an agent x, and an action a just in case p explains why if x does a, then x does something that is intrinsically reason-supported. So on this view, once we get to the ultimate reason-supported actions, there do not need to be any considerations which actually count in favor of doing these things. For an interesting historical view with exactly these commitments, see Nagel [1970]'s introduction of the distinction between objective and subjective reasons. And for discussion, see Chapter 2, this volume.

move their arm in that way, they will perform an action of a type that is intrinsically wrong. And so she must accept that necessarily, if moving your arm in a certain way would torture puppies, then if you move your arm in that way, you will do something that is intrinsically wrong. This conditional is necessary, and appears to connect the pure moral with the non-moral. So it constitutes, I think, the best case to be made that the Pricean strategy is still committed to a violation of Hume's Dictum. But in response, the Pricean can claim that the consequent of this conditional is really just the bastard claim that it is wrong to move your arm in that way. I'm not quite convinced that this amounts to a satisfying answer to the worry about whether we are still violating Hume's Dictum, so I still have some reservations about whether we've fully confronted that form of the explanatory argument. But it makes sense for the proponent of the Pricean strategy to fall back on the fact that this necessity is really a bastard claim, in her defense.

So to sum up, it is still not entirely clear whether we ultimately have a strategy that appeals to necessities that do not themselves constitute some kind of violation of Hume's Dictum. But what is clear, I think, is that appealing to the intrinsic/instrumental distinction, along with the familiar idea that intrinsic wrongness is necessary, which can be supported by the Standard Model Theory, offers the most promising path for making good on the Pricean strategy for responding to the modal challenges to non-reductive moral realism.

6.7 Where This Leaves Us

What I've shown in this paper is that there is at least one strategy that offers at least some initial promise for responding to the Hume's Dictum form of the explanatory argument from supervenience, *without* taking on substantial commitments either in modal metaphysics or about which kinds of supervenience theses turn out to be true.[26] I haven't shown that adopting Price's commitment to the Standard Model Theory of normative explanations is the only way of making good on this strategy, and so perhaps there could be some less committing way of making good on the kind of ideas I've explored in this paper. But I do think that I've shown that it is non-trivial to carry this strategy out, particularly because it is highly non-trivial to offer a plausible rationale for why the moral necessities on which the view is based should intuitively count as *pure*. It is to this question that I think the full set of Price's commitments offer us the most promise of a satisfying answer, for reasons that I've tried to make clear in my discussion of Scanlon.

So what remains of the explanatory argument from supervenience, given the Pricean response? The answer, of course, is—at a mimumum—its second form—the

[26] In contrast, as Schmitt and Schroeder show, the strategy advocated by Wedgwood [1999], [2007] carries not only both of these commitments, but also commitments about the interrelated status of the normative and the intentional.

less direct argument based on the principle that necessities require explanation. This is because the Pricean strategy relies throughout on postulated moral necessities, of which it offers no explanation. Indeed, if the Pricean strategy is really best motivated by the idea I've traced to Cudworth that all moral explanations are subsumptive, then the strategy requires moral necessities that admit of *no* explanation.

The second form of the explanatory argument from supervenience is, of course, less exciting than the form which relies on Hume's Dictum. It gives us no decisive refutation of non-reductive realism; it only probes a source of concern. But the foregoing reasoning leads me to think that this form of the argument is really where the action is. The challenge for non-reductivists is to say why pure moral necessities require no explanation.

One thing that is certainly achieved by the broadly Pricean strategy of responding to the Hume's Dictum argument, is that it forces the non-reductive realist to adopt some clear commitments about the nature of at least some of the pure moral necessities. According to the picture I sketched in the last section, the basic necessities are atomic claims of the form that some action-type has the property of intrinsic wrongness. And according to Scanlon's implementation, the basic necessities are again atomic claims, this time of the form R(p,x,c,a).

We may appeal to what we know about these basic necessities, in order to press the second form of the explanatory argument. For though the relationship between pure and bastard moral claims according to the Pricean strategy is closely analogous to the relationship between pure and bastard mathematical claims, as we've seen, the *kind* of claims postulated to be necessities is actually quite different. The necessities of pure morality, according to the Pricean strategy that I've described, ascribe moral properties to action-types individuated in non-moral terms. Many of them have a simple, atomic structure, ascribing a moral property to a non-moral action-type. The necessities of pure mathematics aren't like this at all. They ascribe mathematical properties—or identity—to mathematical entities. Maybe—just maybe—this makes unexplained moral necessities fishier than unexplained mathematical necessities. Maybe it makes mathematical necessities easier to explain. A full elaboration of the explanatory argument from supervenience would need to say why this would be, and a full proper defense of non-reductive moral realism would need to say why not.

Either way, Price reminds us that the explanatory challenges facing the modal commitments of non-reductive moral realism are complex, and will reward further attention. If modal truths require explanation, then supervenience will not come without some sort of price.[27]

[27] Special thanks to Krister Bykvist, Bart Streumer, John Broome, Shieva Kleinschmidt, John Hawthorne, Jim Van Cleve, Ralph Wedgwood, Jamie Dreier, Geoff Sayre-McCord, Walter Sinnott-Armstrong, Brad Hooker, and audiences at the Carolina Metaethics Workshop in Chapel Hill in December 2012, at the Oxford University Moral Philosophy Seminar and the University of Reading in February 2013, and at Princeton, Chicago, and Northern Illinois University in fall 2013.

Part 3

By contrast, 'How is the imperative of morality possible?' is beyond all doubt the one question in need of a solution. For the moral imperative is in no way hypothetical, and consequently the objective necessity, which it affirms, cannot be supported by any presupposition, as was the case with hypothetical imperatives.

[Kant 2002, 220 (4:419)]

7

The Scope of Instrumental Reason

Allow me to rehearse a familiar scenario. We all know that which ends you have has something to do with what you ought to do. If Ronnie is keen on dancing but Bradley can't stand it, then the fact that there will be dancing at the party tonight affects what Ronnie and Bradley ought to do in different ways. In short, **(HI)** you ought, if you have the end, to take the means. But now trouble looms: what if you have dreadful, murderous ends? Ought you to take the means to them? Seemingly not. But fortunately, an assumption made by deontic logics[1] comes to the rescue. Since "ought", according to this assumption, is a sentential operator, HI must really be *ambiguous*. It could be read either as **(Narrow)** You have the end → O(you take the means) or as **(Wide)** O(you have the end → you take the means). Now if Narrow is true, then you really ought to take the means to your murderous ends. But this doesn't follow from Wide. All that follows from Wide is that you ought to either take the means to these ends or else give them up. Conclusions: (1) Since HI is on some reading true, but Narrow isn't, Wide is true. (2) Wide accounts for the relationship between your ends and what you ought to do.

This elegant scenario repeats itself in many other domains in which it seems like something can have a *bearing* on what some particular agent ought to do. Does what you know affect what you ought to do? Do your beliefs about what you ought to do affect what you ought to do? Do your promises affect what you ought to do? Do your beliefs affect what you ought to believe? On each of these counts, the intuitive answer is "yes". And so each of these questions leaves something for the moral philosopher or the epistemologist to investigate. On each count, it seems that what we all know, is that **(Account)** you ought, if *p*, to do *A*. But on each count, the Narrow-scope reading of the "ought" in this claim yields unintuitive consequences. So since Account is true, it must be true on the Wide-scope reading. So the Wide-scope principle must be what accounts for what each of these things has to do with what you ought to do. The Wide-scope program in the theory of practical and theoretical rationality is to offer

[1] Strictly speaking, this assumption is made only by certain interpretations of deontic logics. On some interpretations, deontic logics only study the logic of "it ought to be the case that *p*," which need not be assumed to have any connection whatsoever to the "ought" of "you ought to go to the store."

these kinds of account, in answer to questions of the form, "what does this or that have to do with what you ought to do?"

Proponents of Wide-scoping[2] hold that this motivation is conclusive. Indeed, they say that Wide-scope principles are "uncontroversial"[3] and that those who do not agree are "confused". It would be nice if this were so. But unfortunately the issues are in fact somewhat more complicated. In this paper, I will explain why the Wide-scoping program is in fact highly controversial. Just as the Narrow principles yield unintuitive results, their Wide counterparts invoke two different kinds of highly controversial commitment. Indeed, I'll argue, on an acceptable semantics for "ought", HI and its counterparts in the other domains are *not*, after all, ambiguous between Narrow and Wide readings. This means that if the Narrow principles are false, so are HI and its counterparts. If this is right, then the Wide-scope views aren't so much offering an *interpretation* of the uncontroversial HI, but are rather *replacing* HI with an alternative that is weaker in one respect, in the face of counterexamples.

Once we see that the project is that of *weakening* HI, however, rather than that of *interpreting* it, I'll suggest that this approach is narrow-minded. If we take a more broad-minded approach, we can see that another way of weakening HI is possible. Instead of weakening it by looking for wider *scopes* for the "ought", we can weaken it by replacing the "ought" itself with a weaker normative concept, such as that of a *reason*. I'll suggest that this is a perfectly viable kind of response to the counterexamples to HI and its analogues. Since this kind of response also escapes the highly controversial features of the Wide-scope accounts, I hold that these kinds of account are actually to be *preferred*. And this, as it turns out, has implications for at least one fundamental and hotly-debated issue in moral philosophy.

7.1 Objective Instrumental Rationality

When Ronnie is keen on dancing and Bradley can't stand it, the fact that there will be dancing at the party affects what each ought to do differently. So being keen on dancing must somehow relate to what one ought to do. But how? That is the question to be answered by an account of objective instrumental rationality. It is an account of instrumental rationality, because it tells us what your ends or desires have to do with what you ought to do. It is an account of *objective* instrumental rationality, because it has to

[2] I don't know how early the Wide-scope view was first formulated, but it is possible that Sidgwick held the view: "When (e.g.) a physician says, 'If you wish to be healthy you ought to rise early', this is not the same thing as saying 'early rising is an indispensable condition of the attainment of health.' . . . [I]t is not merely this relation of facts that the word 'ought' imparts: it also implies *the unreasonableness of adopting an end, and refusing to adopt the means indispensable to its attainment*" (1981, p I iii 4, italics added). More recently, it has been offered as obvious or defended in, for example, Hare [1971], Hill [1973], Greenspan [1975], Darwall [1983], Gensler [1985], Hampton [1998], Broome [1999], Dancy [2000], and Wallace [2001].

[3] According to Stephen Darwall, the Wide-scope view is "uncontestable" [1983, 15]. According to John Broome, it is "an elementary and widely recognized point, but also one that is widely ignored" [1999, 410]. Broome calls disagreeing with Wide-scope views a "confusion".

do with what actions are *actually* means to your ends, rather than with what actions you merely *believe* to be means to your ends. That question is answered by an account of *subjective* instrumental rationality, and we'll return to it in section 7.2.

According to a naïve view, we *already* know at least a *little* bit about how to account for the domain of objective instrumental rationality. It is that HI is true. You ought, if you have the end, to take the means. As noted, if we assume that "ought" takes propositions for one of its relata, then this can receive (at least[4]) two readings:

Narrow ObjO[5]: If you desire[6] that *p*, and your doing *A* is necessary for *p*, then O(you do *A*)
Wide ObjO: O(If you desire that *p*, and your doing *A* is necessary for *p*, then you do *A*)

If Narrow ObjO is true, then from the assumption that you desire to be a successful axe-murderer, and that this requires swinging an axe through someone's body, we can conclude that you ought to swing an axe through someone's body. Surely this is not the case, so surely Narrow ObjO is false.

Wide-scopers therefore conclude that Wide ObjO is true. This follows from the assumption that HI is true, and that Narrow ObjO and Wide ObjO are the two ways of reading HI. I'll now argue, however, that Wide ObjO has two very controversial features. One of these is an outright unintuitive consequence; the other is a feature that many philosophers would be willing to accept, but is still highly controversial. After that, we'll take a closer look at whether HI really is ambiguous in the required way, in the first place.

7.1.1 Symmetry

One difference between Narrow ObjO and Wide ObjO that should be immediately obvious is that Wide ObjO has a certain kind of *symmetry* that Narrow ObjO does not. Narrow ObjO says that if you desire that *p*, and your doing *A* is necessary for *p*, then you ought to do *A*. But it does not say that if you do not do *A*, and you desire that *p*, then you ought to make sure that your doing *A* is not necessary for *p*. Nor does it say that if you do not do *A*, and your doing *A* is necessary for *p*, then you ought to not desire that *p*. Wide ObjO, on the other hand, does posit a symmetry between any of these three ways of complying with its requirement. When Ronnie finds himself desiring to go dancing, and the party is the only place where there will be dancing, Ronnie can satisfy

[4] Once we start having conditionals with conjunctive antecedents, we might think that we should be able to get readings for each combination of the conjuncts outside of the scope of the "ought". For example, here we might expect to get "If you desire that *p*, then O(If your doing *A* can bring about *p*, then you do *A*)" and "If your doing *A* can bring about *p*, then O(If you desire that *p* then you do *A*)." If we really think that "ought" works in this way, then it looks like all of these should be candidates that we should have to worry about.

[5] My convention from here on will be to name principles by whether they are Wide or Narrow, an abbreviation for the domain for which they are supposed to provide an account, and a letter for the normative concept that they employ. So: "O" for *ought*, "R" for *reason*, and so on.

[6] I'm going to conduct the discussion from here on as if the question is what *desires* have to do with what you ought to do. If you think the answer is "nothing", then please feel free to substitute "end" or "intention" wherever appropriate.

Wide ObjO by going to the party. But he can also satisfy it by ceasing to desire to dance. And he can even satisfy it—this is the crazy part—by convincing the party-throwers not to have dancing after all. For if they cancel the dancing, then going to the party won't be necessary for dancing.

Leave aside whether it is rational for Ronnie to react to his situation by ceasing to desire to dance. It *is* the right kind of thing to be a distinctively rational response to his situation, for him to go to the party. But unfortunately, convincing the party-throwers to cancel the dancing does *not* seem to be a distinctively rational response to Ronnie's situation. Whatever else we might say about it, it would be particularly odd for Ronnie, who sincerely desires to dance, to start trying to convince the party-throwers to cancel the dancing, on the grounds the he won't be able to make it. This just doesn't seem like the kind of thing that a good account of objective instrumental rationality should endorse. It is a symmetry predicted by the Wide-scope account that is simply not sustained. Since Wide ObjO makes this prediction, it is clearly problematic.

Bizarrely, Jonathan Dancy has argued that symmetry creates a problem for the *Narrow*-scope view.[7] His argument has three parts. First, he assumes that everyone agrees that Wide ObjO is true. Then, he assumes that the only reason that anyone would believe Narrow ObjO is if it was a consequence of Wide ObjO. And then he uses the symmetry of Wide ObjO in order to derive unintuitive consequences of such a view. Dancy's argument shows that we should not think that Narrow ObjO is a *consequence* of Wide ObjO. But it does nothing to show that Narrow ObjO is itself problematic. For Dancy's argument *against* the Narrow-scope view works by attributing to it commitment to the Wide-scope view as well. Then he uses the fact that commitment to the Wide-scope view engenders a symmetry, to show how this symmetry yields odd results when combined with the Narrow-scope view. But obviously the symmetry only comes in when we accept Wide ObjO. Narrow ObjO by itself has no worries about symmetry.

7.1.2 Agent-Neutrality

Another obvious difference between Narrow ObjO and Wide ObjO is that Wide ObjO is committed to an eternal, agent-neutral obligation, while Narrow ObjO is not. To see how, let's dispense with talk about what "you" ought to do, and state Narrow and Wide in their full quantified glory:[8]

Narrow: $\forall x$ If $p(x)$, then O_x(x does A)
Wide: $\forall x\, O_x$(If $p(x)$, then x does A)

[7] See Dancy [2000, 70–76].

[8] In stating these principles, I am throughout following Broome in assuming that "ought" expresses a *relation* between an agent and a proposition. So I assume that the "O" has always an implicit index. In order to state the quantified version of the principles, however, we need to make the index explicit, so that's what I've done here. See Broome [1999, 399].

Wide tells us that there is something that everyone ought to do. Narrow tells us no such thing. As far as Narrow is concerned, there are only things that particular people ought to do—those people who satisfy the relevant conditions.

This difference between Wide and Narrow lies at the heart of a great controversy about the priority of agent-neutral and agent-relative obligations. Some hold that every time some individual ought to do something, it must be because there is something that everyone ought to do. These are the Neutral-Prioritists. But others reject this. They hold that when there is something that everyone ought to do, that is simply because *each* individual ought to do it. Indeed, the standard definition of agent-neutrality works in this way. It says that there is an agent-neutral reason to do something, just in case there is a reason for everyone to do it. Those who take this line are the Relative-Prioritists.

The divide between Neutral-Prioritists and Relative-Prioritists is old and deep. Neutral-Prioritists are happy to accept principles like Wide. For they think that everything that someone ought to do has to be explained by something that everyone ought to do. For them, the question of how to account for objective instrumental rationality is precisely the question of how to use an agent-neutral obligation, in order to explain how desires or ends can affect what some particular individual agent-relatively ought to do. But Relative-Prioritists are not happy to accept principles like Wide. They hold that not all agent-relative "oughts" can be explained by agent-neutral ones. And they hold that agent-neutral "oughts" carry an explanatory burden. For something must explain why it is that each and every possible agent happens to bear the "ought" relation to this one particular thing. So Relative-Prioritists find principles like Wide particularly suspicious.[9]

One apparently very common kind of Relative-Prioritist view is clearly committed to Wide ObjO being false. This is the view variously known as the "Humean" Theory of Reasons, or the "Desire-Dependence" view. According to this view, all *oughts* or reasons are just like the ones accounted for by the account of objective instrumental rationality. Whenever there is a reason for someone to do something, on this view, it is because doing so promotes one of her desires. This view is not committed to holding that there are no agent-neutral obligations—it is simply committed to holding that if there are, it is *because* they are obligations that happen to be obligations for each agent, rather than vice-versa. This makes it a Relative-Prioritist view. According to E. J. Bond, this view was in fact "the favoured view among professional philosophers" as recently as 1983;[10] T. M. Scanlon recently writes[11] that desires are still "commonly understood" to be the sole source of reasons in this way.

There are actually *two* reasons why Wide ObjO is inconsistent with the "Humean" Theory of Reasons, but one arises simply from the fact that Wide ObjO is committed to explaining Ronnie's case by means of a further, agent-neutral requirement. According

[9] For further discussion, see Chapter 2, this volume.
[10] Bond [1983, 3].
[11] Scanlon [1998, 37].

to the "Humean" Theory, all obligations or reasons get explained in the same way as Ronnie's reason to go to the party gets explained—by desires. But according to the Wide-scope account, Ronnie's reason to go to the party needs to be explained by the existence of a further agent-neutral requirement. This futher requirement therefore can't be explained in the same way as Ronnie's reason to go to the party, because then it would have to be used to explain itself, and that would be circular. So it *can't* be explained by a desire. Anyone who accepts Wide ObjO as an account of objective instrumental rationality, therefore, thereby rejects the "Humean" Theory of Reasons.

It has been argued on these grounds that the "Humean" Theory of Reasons is incoherent.[12] But that's silly. Such an argument employs a controversial premise—the Wide-scope account of objective instrumental rationality. It *may* be that the Wide-scope account is well-motivated. But at the worst, that would pose a dilemma for the "Humean" Theory of Reasons—not demonstrate it to be literally incoherent. Now it may be that the "Humean" Theory of Reasons is false. It may even be, although having thought about the matter a great deal and being consequently sympathetic to the "Humean" theory[13] I would be quite surprised, that it is *obviously* false. But if the "Humean" theory is really "the favoured view" or even "commonly" accepted, the Wide-scope account of objective instrumental rationality simply can't be "uncontroversial".

7.1.3 A Tangent—The Ambiguity in HI

I now turn to whether Wide ObjO is, in fact, well-motivated. Wide-scopers typically motivate Wide ObjO by an argument from elimination. Narrow ObjO yields absurd results, so it can't be true. Therefore Wide ObjO is. But curiously, only two things made it into this argument by elimination. The justification for this is that we already know *something* about how to account for objective instrumental rationality—it is by HI. And the Wide-scoper claims that fortunately, HI is ambiguous between the Wide and Narrow readings. Unfortunately, however, if we take this ambiguity claim seriously, it is rather implausible. For it relies on a very problematic semantics for "ought".

The assumption that we need, in order to get the Wide-scoper's argument going, is that "ought" expresses a relation that takes propositions for one of its objects. The ambiguity proposed by the Wide-scoper is that in HI, the "ought" can be read as taking scope over the whole conditional, or merely over the consequent of the conditional. These are the two sentential clauses in which it figures, so if we think that "ought" takes propositions and works like a sentential operator, then this ambiguity makes sense.

On the face of it, however, "ought" does not take propositions. It takes *actions*, in some very broad sense—things that people can *do*. This is why "there is something that

[12] On a very natural reading, this is the central argument of Jean Hampton in *The Authority of Reason*. Hampton, however, isn't particularly clear on this point. I can generate two other, quite different readings, of what her central argument might be. See Hampton [1998].

[13] I defend a version of the "Humean" theory in Schroeder [2007a].

you ought to do" follows from "you ought to go to the store" and "going to the store is something that you ought to do" rearranges pleonastically with "you ought to go to the store." Propositions are not things that you can *do*. So if "ought" takes propositions, then "there is something that you ought to do" should not follow from "you ought to go to the store." Something like "there is something that you ought to make true" should follow instead. Likewise, *going to the store* is not a proposition. It is an action-type. So if "ought" takes propositions, then it is hard to see why "going to the store is something that you ought to do" should pleonastically rearrange with "you ought to go to the store." But if "ought" takes action-types, on the other hand, then this is easy to see. On this view, these two sentences are related to one another in the same way as "Mary is left of John" is related to "John is someone Mary is left of."

The thesis that "ought" takes propositions, as John Broome notes,[14] is not without linguistic evidence. The evidence for this view comes from an attractive proposal for how to understand infinitive clauses like "to go to the store." Compare "he wants to see the Pacific" to "he wants her to see the Pacific." On a natural view, these two sentences should be accounted for along similar lines. So on a natural view, there must be a hidden pronoun in "he wants to see the Pacific." It must really be a little bit like, "he wants himself to see the Pacific." Granting the existence of such hidden pronouns, it looks like the infinitive clause gives us something very proposition-like. On the other hand, for the reasons just cited, and a few others, we shouldn't get over-excited by this kind of evidence. Compare these sentences to one like "it is wrong to murder children." What is the hidden pronoun in "it is wrong to murder children"? For whom does it say that murdering children is wrong? Perhaps such sentences are best treated as involving some kind of generic or universal quantifier, but this is surely highly controversial, at best.[15] This sentence seems to pleonastically rearrange with "murdering children is wrong" and to predicate wrongness of an *action*—murdering children. Whatever kind of things we *ought* to do, plausibly they are the same kind of thing as whatever kind of things are *wrong*.

Worst of all, the thesis that "ought" takes propositions yields some intolerable predictions. It predicts that it should be at least conceptually possible that you ought that *I* go to the store. After all there is a proposition that I go to the store. And you are an agent. And those are the kinds of thing that the "ought" relation holds between. So all it takes for it to be the case that you ought that I go to the store is that you stand in the *ought* relation to this proposition. That is, that O_{you}(I go to the store). Now, it is certainly possible that you ought to make sure that I go to the store. And it is certainly possible that you ought to tell me to go to the store, and that you ought to *help* me go to the store. These are the possibilities that O_{you}(you make sure that I go to the store) and O_{you}(you tell me to go to the store) and that O_{you}(you help me go to the store). These

[14] As Broome notes [1999].

[15] See Chapter 2, this volume, for further discussion of the philosophical (not semantic) reasons why this treatment of "wrong" should be so highly controversial.

possibilities all make sense. But if Broome's view about "ought" is right, then there should be another possibility, distinct from all of these: the possibility that O_{you}(I go to the store). Frankly, I can't see what this could even *be*. It sounds like a category mistake.

Broome is happy to bite this bullet. In print and in personal conversation, he has expressed his regrets that English grammar does not let us talk about such interesting possibilities as your *ought*-ing that I go to the store.[16] But that *is* simply bullet-biting. There is no such possibility for us to talk about. The claim that you ought that I go to the store isn't simply an ungrammaticality. After all, the claim that you ought that *you* go to the store is also ungrammatical, but we can at least understand what it means. The claim that you ought that I go to the store is worse than ungrammatical, for so long as we distinguish it from each of the other things we distinguished it from above, none of us have any idea what it means. It patently manifests a category mistake. So "ought" simply can't properly express a relation between agents and propositions. If it did, there really would be such possibilities to talk about.

I've been following Broome in taking the reasonable view that "ought" takes agents for one of its relata. But there is another view on which "ought" takes propositions instead of actions, but on which it is not a relation between an *agent* and a proposition. It treats "ought" like the English expression, "it ought to be the case that," as expressing a monadic property of propositions. Those who hold this view have a different but related problem to deal with. They must give us an analysis of "*you* ought to go to the store." On the natural version of this view, for it to be that you ought to go to the store is just for it to be the case that O(you go to the store). But there is an important difference between "you ought to go to the store" and "the deficit ought to shrink" that this analysis seems not to capture. For you, unlike the deficit, are an *agent*. And most philosophers think that there is a sense in which it can be that an agent ought to do something that cannot apply to non-agents. But the deficit figures in the subject-place of O(the deficit shrinks). So it looks like if "you ought to go to the store" follows from O(you go to the store), then "the deficit ought to shrink" must follow in precisely the same sense from O(the deficit shrinks). So this view, like Broome's, has obvious troubles making the right predictions about the sense in which an agent ought to do something. Broome's view predicts too many things for it to be that an agent ought to do. This other view predicts too many things to qualify as "ought"-ing to do something.[17]

7.1.4 Reasons

If "ought" does not take propositions, then HI is not, after all, ambiguous between Narrow and Wide readings. HI is a conditional, conditionals are not actions, and the

[16] Broome [1999], and also in personal conversation.

[17] Nevertheless, I am going to continue to use Broome's notation in stating the Wide-scope views. Many (but not all) of the Wide-scope views can plausibly be intelligibly restated without the assumption that "ought" takes propositions. By rejecting this assumption, we simply lose the *motivation* for thinking that we are somehow *preserving* the truth of HI. And so we lose the motivation for not including other kinds of account in the argument by elimination for the Wide-scope view.

Wide-scope reading requires that the "ought" take the entire conditional for its scope. So Wide ObjO is not an admissible disambiguation of HI. This means that if Narrow ObjO is false, then HI is false. And if HI is false, then the right way to give an account of objective instrumental rationality must be to discover what is true *instead* of HI—not to find a reading of HI on which it is true.

Now, this is something that we can understand Wide-scopers as trying to do. On this revised reading, Wide-scopers are not offering Wide ObjO as a *disambiguation* of HI. They are offering it as a *replacement* for HI—that is, for Narrow ObjO. It is weaker than Narrow ObjO in one relevant respect. From Wide ObjO and the assumption that you desire that *p* and your doing *A* is necessary for *p*, nothing follows about whether you ought to do *A*. That such conclusions *did* follow from Narrow ObjO was precisely what was wrong with it.

But now that we are engaged in this project, it is easy to see that there are other ways of weakening Wide ObjO in order to get this result. For example, we can replace talk about what you *ought* to do with talk about what there is a *reason* for you to do. We can assume that if you ought to do something, then there must be *some* reason for you to do it. But you have reasons to do many things that you ought not to do—even things that you *patently* ought not to do. For not all of your reasons are very good. The reasons for you to perform some particular action are a little bit like the items which appear in the "pros" column when God sits down and lists all of the pros and cons of your performing that particular action, with a view to advising you about whether to do it. Even if God always advises you one way or the other, he almost always has at least something to mark in each column. So there is almost always at least *some* reason in favor of any course of action, even ones you patently ought not to take.

Now, it is reasonably obvious that even when you desire to become a successful axe-murderer, you still ought not to sharpen your axe or stake out victims, let alone swing your axe at people. But it is less obvious that you have *no* reason whatsoever to do so. After all, the reason might simply not be very good—and we can agree that the reasons for you *not* to do these things are about as excellent as reasons come. So even if you do have some reason to do these things, we need have no worries about whether it will turn out that you ever *ought* to do them. This gives us a quite different, *Narrow*-scope way of weakening Narrow ObjO in order to avoid its unintuitive results:

Narrow ObjR: If you desire that *p*, and your doing *A* promotes *p*, then there is a reason for you to do *A*.

Narrow ObjR is clearly a Narrow-scope account of objective instrumental rationality. But it is not at all obvious that its consequences are intolerable.[18] Notice that in addition to weakening Narrow ObjO by changing "ought" to "reason", I've strengthened it, by replacing "is necessary for *p*" with "promotes *p*." I'll return to discuss this kind of change when we get to the discussion of theoretical rationality, in section 7.5.

[18] See Schroeder [2007c] for an extensive discussion of whether Narrow ObjR is still unacceptably strong.

7.2 Subjective Instrumental Rationality

As it turns out, different issues arise, and with more or less force, when we consider the Wide-scoping program in the different domains in which it is applied. So in the next few sections I'm going to go through and consider each of four more such applications, in order to bring out a few more complications. The case of Ronnie and Bradley clues us in to the fact that we need an account of objective instrumental rationality. Ronnie differs from Bradley because he is keen on dancing, but Bradley is not. Now Freddie, like Ronnie, is keen on dancing. But Freddie knows something that Ronnie does not. He knows that there will be dancing at the party. Just as Ronnie differs from Bradley with respect to whether each ought to go to the party, so also Ronnie differs from Freddie. We'd be surprised if Ronnie went to the party, but not surprised if Freddie went. We'd think Freddie irrational for not going, but not so Ronnie. The difference between Ronnie and Freddie is accounted for by an account of *subjective* instrumental rationality.

Philosophers discussing instrumental reason are often not very careful to explain whether they are talking about the difference between Ronnie and Bradley, or the difference between Ronnie and Freddie. But these *are* two distinct differences, and this is important. A Wide-scoper would have it that we can all agree that **(HI+)** you ought, if you desire that *p* and believe that your doing *A* is necessary for *p*, to do *A*. But of course this gets two readings:

Narrow SubjO: If you desire that *p*, and you believe that your doing *A* is necessary for *p*, then O(you do *A*)
Wide SubjO: O(If you desire that *p*, and you believe that your doing *A* is necessary for *p*, then you do *A*)

Narrow SubjO is twice as unintuitive as Narrow ObjO. For now there are two ways to derive crazy results from Narrow SubjO. We can assume that you have crazy desires, or we can assume that you have crazy beliefs about how to accomplish your desires. For example, you might desire to succeed in your career, and falsely believe that murdering me in cold blood and spreading my remains around your boss's office is the way to do so. But it hardly seems to follow from this that you ought to murder me in cold blood and spread my remains around your boss's office. So Wide SubjO is to be preferred to Narrow SubjO.

7.2.1 Symmetry

Like all Wide-scope principles, Wide SubjO posits a symmetry between different ways in which it might be fulfilled. Freddie can satisfy the requirement posed by Wide SubjO in any of three ways. He can go to the party, or he can stop desiring to dance, or he can change his mind about whether there will be dancing at the party. But this is peculiar. Surely, concluding that there will not be dancing at the party after all is not a distinctively instrumentally rational way of responding to Freddie's situation.

We need to be somewhat careful, here. For the Wide-scoper can say that this is indeed an irrational way for Freddie to respond—because it is ruled out by his account of *epistemic* rationality. So we don't get a problem for the Wide-scoper merely by noticing that so far as it says, it may be rational for Freddie to change his belief. The problem for the Wide-scoper is that if Freddie *does* respond to his situation in this way, she has to allow that though Freddie is being *epistemically* irrational, he is in fact behaving impeccably, when it comes to subjective instrumental rationality. And that is a bizarre thing to say. Surely a good account of subjective instrumental rationality should not tell us that so far as instrumental rationality goes, this kind of behavior is okay.

Freddie's case illustrates an important point. Wide-scope principles are good at predicting what is wrong with an agent *at a time*. But they are not good at predicting the rational ways for an agent to *change* her situation. Now, not all domains in which the Wide-scoping program is applied are domains in which we are particularly interested in how it is rational for an agent to respond to her situation. For example, this is explicitly one difference between the domain of objective instrumental rationality and the domain of subjective instrumental rationality. It is *irrational* for Freddie not to go to the party, but not irrational for Ronnie not to go. Since Ronnie doesn't know anything about the party, going there is no more rational than not—even though, in some sense, it is what he ought to do, given his ends.

In Freddie's case, if he is not going to the party, something is going badly. He wants to dance, he believes that he can only dance by going to the party, and he doesn't go. If he then changes his mind about whether there will be dancing at the party, then he puts himself in a better position. He no longer has this kind of inconsistency between his aims, beliefs, and actions. So he takes himself from a worse position to a better. But despite the fact that this kind of move makes him more rational at a time, it is not a rationally permissible move. If we want an account of subjective instrumental rationality to specifically tell us something about what moves it is rational for Freddie to make, then we should be particularly sensitive to the fact that the Wide-scope account predicts only symmetries.

7.2.2 Agent-Neutrality

According to Wide SubjO, there is something that everyone ought to do—to not have desires, beliefs, and actions in conflict with one another in the way that Freddie's are, when he doesn't go to the party. But it's hard to see where the obligation to do this comes from. It doesn't arise because of desires, nor because of beliefs. A Narrow-scope view can say where the obligations or reasons that it posits come from—they arise as a result of beliefs or desires. The Wide-scope view, on the other hand, needs to posit *unexplained* obligations or reasons.

In fact, if we accept Wide ObjO and Wide SubjO, then we have to posit two distinct eternal, agent-neutral requirements in order to explain what is (distinctively) wrong with Freddie when he doesn't go to the party and what is wrong with Ronnie when he

doesn't go. But on a natural view, there should be some common explanation of what goes wrong with Ronnie and what goes wrong with Freddie. They should be related in some intimate way. I'll illustrate how to give such an explanation in the next subsection.

7.2.3 Subjective Reasons

As in the theory of objective instrumental rationality, I'm going to suggest that Wide-scopers weaken Narrow SubjO in the wrong way. Or at least, they don't weaken it in the only plausible way. But in this case, it is not sufficient to say that there is a *reason* for Freddie to go to the party. For though Freddie believes there to be dancing at the party, perhaps it turns out that Freddie is wrong. If there is no dancing at the party, then it follows from the account of objective instrumental rationality that there is a reason for Freddie *not* to go to the party. But it sounds more than odd to say that at least there is this much to be said for his going there anyway: at least he *believes* that there will be dancing there. That shouldn't make it into God's list of pros and cons. So the theory of subjective instrumental rationality raises a new puzzle not raised by the theory of objective instrumental rationality. If we are to weaken Narrow SubjO, we need to weaken it *more* than by changing "ought" to "reason".

I think that this shouldn't be surprising. For I don't think that we should want a distinct account of subjective instrumental rationality in the first place. Compare Ronnie and Freddie to Ryan and Bryan. Katie needs help, and that's a reason to help her. It's a reason for Ryan to help her, and a reason for Bryan to help her. But only Bryan knows that Katie needs help. Ryan is blissfully unaware. So Ryan and Bryan differ in what we can expect of them, in both the predictive and the normative senses. We can expect Bryan to help Katie, but we can't expect any such thing of Ryan. We can blame Bryan for not helping her, but we can't blame Ryan.[19]

It looks like Bryan differs from Ryan in exactly the same way as Freddie differs from Ronnie. There is some reason for each to do something, but only Bryan and Freddie are aware of these reasons. There is a special kind of status that you have, when you believe something that, if it is true, is a reason for you to do something. In ordinary English, in fact, we can even use the word "reason" to describe your situation. Consider the familiar case of Bernie: Bernie is at a cocktail party, holding a glass of gasoline that he believes to be a gin and tonic. Intuitively, the fact that his glass is full of gasoline is a reason for him not to take a sip. But in another sense, this isn't one of his reasons—at least, it isn't a reason that he *has*, since he is unaware of it. In this second, perfectly legitimate sense, Bernie *does* have a reason to take a sip—for he reasonably believes that his glass contains the gin and tonic for which he asked the hostess. Carefully spelled out, Bernie's case gives us cause to distinguish these two senses of the word "reason": call them *objective* and *subjective*.

[19] Here I ignore for simplicity the complication that his ignorance might turn out to be culpable.

On a natural view, subjective reasons are simply things that you believe such that, if they are true, they are reasons for you to do something. That seems to be what is going on with Bernie, it seems to be what is going on with Bryan, and it seems to be what is going on with Freddie. If this is right, then it follows from our account of *objective* instrumental rationality that Freddie has a *subjective* reason to go to the party that Ronnie doesn't have. For Freddie, but not Ronnie, believes that there will be dancing at the party, and this is the *objective* reason for both of them to go there.

Taking this very natural view of the matter commits us to an even weaker account of subjective instrumental rationality:

Narrow SubjSR: If you desire that *p*, and you believe that your doing *A* is necessary for *p*, then you have a subjective reason to do *A*.

But on this view, Narrow SubjSR is not a distinct theoretical posit, needed in order to explain what is going on in the case of subjective instrumental rationality. It falls neatly out of our already-existing Narrow-scope account of objective instrumental rationality, and our very natural account of the relationship between the objective and subjective senses of the word "reason", which we are independently forced to acknowledge.

7.3 The Role of Conscience

Wide-scoping is also commonly employed to explain the relationship between your conscience and what you ought to do. According to the Wide-scoper, we can all agree that in some sense or other **(Consc)** you ought, if you believe that you ought to do something, to do it. You should, that is, let your conscience be your guide. But as always, the Wide-scoper holds that this claim is ambiguous:

Narrow ConscO: If you believe that O(you do *A*), then O(you do *A*)
Wide ConscO: O(If you believe that O(you do *A*), then you do *A*)

Now, Narrow ConscO yields some unfortunate results. From it, it follows that you are infallible with respect to what you ought to do. And surely that is false. Surely you can be mistaken about what you ought to do. So Wide ConscO is definitely to be preferred to Narrow ConscO.

7.3.1 Symmetry

But let us not get ahead of ourselves. Wide ConscO, too, has its problems. For one, there are *two* ways to comply with Wide ConscO. You can comply with it either by doing what you believe you ought to do, or by changing your mind about what you ought to do. But surely there is a relevant asymmetry, here. After all, we have a special name for the distinctive vice of changing your mind about what you ought to do, simply so that you don't have to do it. It is called *rationalization*. The whole point of conscience being your guide is that changing your beliefs about what you ought to do simply in order to avoid doing it is *not* an acceptable way to proceed.

7.3.2 Agent-Neutrality

Like the other Wide-scope accounts, Wide ConscO works by positing a basic, eternal, agent-neutral requirement rationally binding on every agent, no matter what they are like. As with all of the others, it offers no explanation of this requirement. Narrow-scope accounts can explain the obligations or reasons that they postulate. After all, these obligations or reasons only exist given a certain condition—so we can use that condition to explain them.[20] But not so for the Wide-scopers.

7.3.3 Subjective Reasons Again

The agent-neutral requirements postulated by the Wide-scoper involve even more commitment, once we see that the requirements postulated by Wide ObjO, Wide SubjO, and Wide ConscO are all distinct. I propose, however, that we can reject Consc altogether, at least if we understand both "oughts" in the same *objective* sense. And as in the other cases, I suggest that our theory can be better—and more economical—if we find a different way to weaken Consc. As with the theory of subjective instrumental rationality, I suggest that the necessary weakening involves the notion of a *subjective* reason:

Narrow ConscSR: If you believe that O(you do *A*), then you have a subjective reason to do *A*.

On the natural view that I will suggest, Narrow ConscSR simply falls out of the account of what subjective reasons are, from section 7.2. This is because on a natural view, the fact that you ought to do *A* is a reason for you to do *A*.

A number of philosophers have recently rejected this natural view. They claim that the fact that you ought to do *A* only *reports* the existence of *other* reasons for you to do *A*—it is not itself a reason for you to do *A*. Their argument is twofold: first, it can't be the case that you ought to do *A* unless there is some *other* reason for you to do *A*. And second, the fact that you ought to do *A* shouldn't be weighed *separately* from these other reasons, in determining whether you ought to do *A*. It doesn't carry any extra weight. I agree with both of these claims. But weighing reasons is complicated. Perhaps two things can both be reasons, even though they shouldn't be weighed separately. For example, in ordinary English, we can say that the fact that there will be dancing at the party is a reason for Ronnie to go there. But in ordinary English we can also say that the fact that Ronnie is keen on dancing is a reason for Ronnie to go to the party. Surely, if these are both reasons for Ronnie to go to the party, they shouldn't be weighed separately. Now some hold that this is an argument that ordinary English speaks falsely

[20] This is exactly what Kant says about why hypothetical imperatives are easier to understand than categorical imperatives: "On the other hand, the question as to how the imperative of morality is possible is undoubtedly the only one requiring a solution. For it is not at all hypothetical; and hence the objective necessity which it presents cannot be based on any presupposition, as was the case with the hypothetical imperatives." Kant (1997a, 4:419). It is strange, then, that many Wide-scopers interpret Kant as sharing their view about objective instrumental rationality. I argue that Kant is not a Wide-scoper in Chapter 9, this volume.

on these counts, and these are really the same reason. But I hold that we could just as well say that adding up the weights of reasons is more complicated than simply placing weights on a scale. If that is right, then we can say that the fact that you ought to do *A* itself counts as a reason for you to do *A*.

Indeed, given our account of the relationship between objective and subjective reasons, this is precisely what we should say. For consider the case of John.[21] John loves successful surprise parties thrown in his honor, but hates all other parties—most of all unsuccessful surprise parties thrown in his honor. In the next room, all of John's friends are waiting, ready to surprise John with a party of which he so far has no clue. There is an excellent reason for John to go into the next room—that a successful surprise party is waiting for him. But you could never give him this reason, because that would make the reason itself disappear. So instead you merely tell him that he *ought* to go into the next room. Does John have a subjective reason to go into the next room? Well, we can expect him to go, and it would be irrational of him not to go there. If he goes, we won't say that he went for no reason at all—we'll say that he went because he ought to. If this is right, then John has a subjective reason to go into the next room by believing that he ought to go. And so by our account of subjective reasons, the fact that he ought to go into the next room must itself count as an objective reason for him to do so. And granting this, Narrow ConscSR falls out immediately from our account of subjective reasons.

7.4 Promises

A very old application of Wide-scoping is to promises. This domain sheds some light on another possible motivation for Wide-Scoping, as well as on the kind of commitment I've been discussing under the heading of "agent-neutrality". Intuitively, the issue is this: Al promises Rose to meet her for lunch. Fortunately for Al, if something comes up, Rose has the power to excuse him from this promise. But if she doesn't, then something is amiss if Al doesn't show up for lunch. According to the Wide-scoper, we can all agree that in some sense or other **(Promise)** you ought, if you promise *Y* to do *A* and *Y* doesn't excuse you, to do *A*. And as always, the Wide-scoper claims that we can read this in more than one way:

Narrow PromiseO: If you have promised *Y* to do *A*, and *Y* has not excused you from doing *A*, then O(you do *A*)
Wide PromiseO: O(If you have promised *Y* to do *A*, and *Y* has not excused you from doing *A*, then you do *A*)

The normal Wide-scoping puzzle would be that Narrow PromiseO apparently can lead to some funny consequences. What if you promise someone to commit some foul

[21] Thanks to Nate Williams for this case, even though he wouldn't approve of this use of it.

deed? Ought you to commit the foul deed? Or what if you make conflicting promises? Ought you to do both? To avoid these results, the Wide-scoper would have us to prefer Wide PromiseO to Narrow PromiseO.

But in fact, this is not usually what motivates Wide-scoping about promising. What usually motivates Wide-scoping about promising is a much simpler observation: that there is only one situation in which something is really going wrong with Al. It is the situation in which he makes his promise, Rose does not excuse him from it, and he fails to show up for lunch. The possibilities are depicted in the following table:

	Al shows up for lunch.	**Al doesn't show up.**
Doesn't give permission.	Okay	! something amiss!
Rose gives permission.	Okay	Okay

The Wide-scoper's natural idea about Al and Rose's case is that what we need is some way of *ruling out* the situations in which something goes amiss. So that is precisely what she does. She postulates a special requirement that says, "don't be like that," demonstrating the cases in which something goes amiss because an agent makes a promise, is not excused, and doesn't keep it.

This kind of motivation can be supplied for Wide-scoping in other domains. In the theory of objective instrumental rationality, we want to know what is amiss with Ronnie, when he desires to dance, going to the party is necessary for dancing, and he doesn't go to the party. So the Wide-scoper postulates a special requirement that rules out precisely those kinds of situation. In the theory of subjective instrumental rationality, we want to know what is amiss with Freddie, when he desires to dance, believes that there will be dancing at the party, and doesn't go to the party. So the Wide-scoper postulates a special requirement that rules out precisely those kinds of situation. In theorizing about the role of conscience, we want to know what is amiss with you, when you believe that you ought to do *A*, but don't do *A*. So the Wide-scoper postulates a special requirement that rules out precisely those kinds of situation. In this way, Wide-scoping gains incredibly elegant solutions to each of these problems—always by postulation of some new eternal, agent-neutral requirement, irreducible to any of the others. Elegant solutions, at the cost of unexplained, basic, agent-neutral requirements. This is a distinct kind of motivation for Wide-scoping.

7.4.1 Symmetry

The Wide-scope account of Promising postulates a single requirement ruling out precisely those situations in which something is amiss. So to satisfy this requirement, all that Al has to do is to get out of this situation. Given that he's already promised to meet Rose for lunch, he can't change that fact. But he *can* solicit her permission not to show up. Rose has the power to *excuse* Al from showing up for

lunch. If Al convinces her to do so, on the Wide-scope view, he is *satisfying* all of his relevant obligations.

But that is clearly wrong. There is at least one obligation that Al is precisely *not* satisfying by soliciting Rose's permission not to show up for lunch. It is an obligation that he is getting *out* of. And this is the obligation to show up for lunch. Though Rose has the authority to *dismiss* Al's obligation to show up for lunch, that doesn't change the fact that he *does* have such an obligation. It is an obligation that he *satisfies* by showing up for lunch, but merely *escapes* by getting Rose's permission not to. Either way, he is only *violating* the obligation if he both does not show up for lunch and does not get her permission.

The problem is that the table only shows us when Al has fallen astray of some obligation *or other*. It doesn't tell us anything about which, if any, obligations he is *satisfying*, or which, if any, obligations he is *escaping*. Unlike the asymmetries that we've diagnosed so far, this is not an asymmetry that only exists when we look at how we want or expect an agent to behave *over time*. In Al's case, asking Rose's permission not to show up for lunch is in perfect keeping with the rules governing promises. It's a *perfectly* acceptable thing to do. But after he has done it, the correct way to describe his situation is that his obligation has *gone away*—not that he has satisfied it.

7.4.2 Agent-Neutrality

The Wide-scope account of promising therefore deals with Al's case a little bit *too* neatly. The mere fact that something is going amiss in a large class of situations doesn't demand that there be any requirement in particular that rules all of those situations out. In Al's case, Narrow PromiseO succeeds in ruling out all and only the same situations as does Wide PromiseO. But Narrow PromiseO rules out these situations on a case-by-case basis. When Al doesn't get permission from Rose and fails to show up for lunch, the obligation that he is violating is his obligation to meet Rose for lunch. And we know where that came from—he promised her to. Likewise, when Sylvia fails to repay her loan and is unexcused, the obligation she is violating is her obligation to repay her loan. And Narrow Promise tells us where this obligation came from, as well—it arose as a result of the promise that she made when she asked her parents for money.

The promising case is therefore good for illustrating how a Narrow-scope account can rule out all of the same situations as a Wide-scope account, but without positing symmetry, and without committing to a single and unexplained requirement that rules those situations out specifically. And this further illustrates the nature of the commitment of Wide-scopers that I've been discussing under the heading of "agent-neutrality".

7.4.3 Obligations

In the case of promising, I don't think that much needs to be done in order to amend Narrow PromiseO. This is because the consequences of Narrow PromiseO, interpreted

correctly, are not particularly unintuitive. In fact, in this case we know from the asymmetry of satisfying obligations and escaping them that we *need* at least one Narrow-scope principle in order to distinguish the case in which Al gets excused from that in which he keeps his promise. And once we have that principle, any Wide-scope account would be superfluous, and can be done without.

Still, philosophers are often picky about the word "ought". It is often held that it can't be the case that you ought to do *A* and that you ought to do *B*, if doing *A* and doing *B* conflict. This view follows, in fact, from the principle that *oughts* aggregate across conjunction, and the common view that "ought" implies "can". But it is clearly possible to make conflicting promises. I'm not sure what to think of either of these two principles, but if this is really how "ought" works, then fortunately there is a very simple way of weakening Narrow PromiseO to avoid this result:

Narrow PromiseOb: If you have promised *Y* to do *A*, and *Y* has not excused you from doing *A*, then you are under an obligation to do *A*.

Narrow PromiseOb posits an obligation for you to do *A*, when you promise to do *A*. On this view, you may have conflicting obligations, but it does not follow from the fact that you are under an obligation to do *A*, that you ought to do *A*. This only follows if you are under no conflicting obligations. So it still turns out that it can't be the case that you ought to do *A* and that you ought to do *B*, if doing *A* and doing *B* conflict.

7.5 Epistemic Rationality

Philosophers defending Wide-scope accounts of one or another of the domains discussed above often appeal to the domain of epistemic rationality as a case in which Wide-scoping obviously applies, in order to gain credibility for their views. Intuitively, the case is this: Phil believes that *p*, and that if *p* then *q*. But he also disbelieves that *q*. Patently something is amiss with him. At least four different kinds of account might be offered:

Narrow Bf(wk)O: If you believe that *p* and that if *p* then *q*, then O(you don't disbelieve that *q*)
Wide Bf(wk)O: O(if you believe that *p* and you believe that if *p* then *q*, then you don't disbelieve that *q*)
Narrow Bf(sg)O: If you believe that *p* and that if *p* then *q*, then O(you believe that *q*)
Wide Bf(sg)O: O(if you believe that *p* and you believe that if *p* then *q*, then you believe that *q*)

The weak (wk) accounts govern only whether you should disbelieve that *q*. The strong (sg) accounts govern whether you should actually *believe* that *q*. Here I assume that one can withhold belief from a proposition—take no opinion about it. This is neither believing nor disbelieving.

Wide Bf(wk)O successfully rules out the worst set of cases: those in which you actually have contradictory beliefs. Something is amiss with you if you have contradictory beliefs, whether or not you realize it. But Wide Bf(wk)O doesn't tell us very much about how it is

rational for you to react to your situation. For it doesn't even rule out the situation in which Phil believes that *p*, believes that if *p* then *q*, is actively wondering whether *q*, and still has no opinion about whether *q*. And that, surely, is a situation in which it is not epistemically rational for Phil to find himself.

Yet Wide Bf(sg)O seems to rule out too much. For it rules out situations in which Phil believes that *p* and believes that if *p* then *q* but has never put these two thoughts together, which is what explains why he has never formed an opinion about whether *q*. And that is clearly ruling out too much. It is not epistemically irrational at all to be like that. Nor is it rationally required for you to have a deductively closed set of beliefs.

7.5.1 Symmetry

Wide Bf(sg)O also predicts a tolerable, if slightly surprising, symmetry. It predicts that if you find yourself believing that *p* and that if *p* then *q* and having no opinion whatsoever about *q*, one epistemically rational way for you to respond to your situation is to cease to believe that *p*. This is at least initially somewhat surprising. But I don't think that it is intolerable.[22] A worse symmetry prediction comes to light when we try to repair Wide Bf(sg)O in order to solve the problem just posed for it in the last subsection.

The problem is that Wide Bf(sg)O tells us that *too* much is wrong with Phil—it tells us that something would be wrong with Phil even if he wasn't bearing his beliefs fully in mind or wondering whether *q*. So on the face of it, what we want is to somehow incorporate the facts that Phil is wondering whether *q* and that he is paying attention to his relevant beliefs. The Wide-scope way to do this is simple:[23]

Wide Bf(sg)O*: O(if you believe that *p* and you believe that if *p* then *q* and you are paying attention to these beliefs and you are wondering whether *q*, then you believe that *q*)

But there are *five* ways to comply with Wide Bf(sg)O*. For example, Phil can comply with it by ceasing to wonder whether *q* or by ceasing to pay attention to his relevant

[22] It *may* be, for example, that the reason that you withhold belief from *q* is not that you have no evidence either way, but rather that you have *too much* evidence *both* ways, and so you can't decide the issue. If you have enough evidence both ways, you might think that no evidence that *p* could possibly be robust enough to decide the issue whether *q*. And so the evidence that ~*q* might therefore be sufficient to override your evidence for believing that *p* even without being sufficient to justify your believing ~*q*. This seems like a perfectly coherent scenario to me, so perhaps this symmetry prediction is quite alright, initially surprising as it might be.

[23] There is a natural way of trying to give a *mixed* wide/narrow account in order to fix this problem:

> **Mixed Bf(sq)O:** If you are paying attention to your belief that *p* and your belief that if *p* then *q*, and you are wondering whether *q*, then O(if you believe that *p* and you believe that if *p* then *q*, then you believe that *q*)

The mixed account is Narrow-scope with respect to the paying-attention and wondering, but Wide-scope with respect to the believing. But unfortunately, it is hard to make sense of the mixed view in a way that respects the fact that it aspires to be Wide-scope with respect to the believing. For paying attention to a belief isn't independent from having that belief—it seems to *require* having that belief. So once we move these conditions outside of the scope of the "ought", we've already moved the believing condition outside of the scope of the "ought", and then we don't have a mixed view at all, but only a Narrow-scope view.

beliefs. Now, it may be that, considered at a time, Phil is being more rational to still wonder whether *q* if he is not paying explicit attention to the beliefs that entail it. So statically speaking, ceasing to pay attention to these beliefs can take Phil from a position where more is amiss with him, to a position in which less is amiss with him. It can make him better off, rationally speaking, if we are only making rational assessments of him at particular times. But it hardly follows that this is a rational thing for Phil to do! On the contrary, this looks like a paradigm of epistemic irrationality. So Wide Bf(sg)O* makes a prediction of symmetry that seems to be unfulfilled. Or, put differently, this is a respect in which the domain of theoretical rationality calls for an *asymmetry* for which the Wide-scope account, by itself, is unable to account.

The symmetry problems for the Wide-scope account of epistemic rationality get much worse, when we try to expand the account, in order to deal with cases of non-deductive inference. On the face of it, epistemic rationality does not only have something to do with how we deal with deductively valid arguments. It also has something to do with how we treat inductive evidence. But Wide-scope accounts are designed merely around this special case. They become much more problematic when we try to expand our account in order to deal with inductive evidence.

Suppose, then, that you believe that Carrie is 99% reliable, and that Carrie said that *p*, but are undecided whether *p*. If you have no other reason to believe that ~*p*, there is something odd about you not going on to believe that *p*, in this case. There is no *strict* requirement here, of course, even when you are bearing both of these supporting propositions fully in mind. But most belief formation is not deductively valid. It is based on evidence that is less than fully conclusive, but that doesn't mean that there aren't any questions about which ways of proceeding are epistemically rational, and which are not.

The Wide-scoper about epistemic rationality cannot simply say that there is a strict requirement that you not be in all three of these states at the same time. For once the evidence is less than conclusive, the requirement must also be less than strict. Although it makes sense to conclude that *p* on the basis of the belief that Carrie is 99% reliable and that Carrie said that *p*, you can in all rationality believe these things but believe that ~*p*—for example, if you saw that ~*p* with your own eyes. If the Wide-scoping program is to have any general applicability, therefore, the "requirement" which forbids being in all three of these states at the same time must be less than strict—it must be merely one of degree.

Imagine, then, that the Wide-scoper tells us that when a set of beliefs is *unlikely* to be true together, there is a *reason* not to believe all three of them which varies in strength according to *how likely* they are not to be all true together.[24] This is an unpromising tactic, since it makes symmetric predictions about the rationality of giving up each belief. But in fact, "Carrie is 99% reliable" and "Carrie said that *p*" are much *better* evidence that *p* than "~*p*" and "Carrie said that *p*" are for "Carrie isn't 99% reliable." Posing only

[24] This, I assume, is the obvious way for the Wide-scoper to tackle this issue. I don't assume that it's the *only* way for such a view to tackle the issue, for it's not my purpose to refute the Wide-scope view. I'm only trying to demonstrate that such a view has some hard questions to answer, before we can agree that it is an *adequate* account of the domain, let alone that it is uncontroversial.

a slack requirement not to believe all three therefore loses track of this important element of the structure of the relationship between these three propositions.

If Wide-scoping is not the right way to account for *inductive* applications of epistemic rationality, however, then it is highly implausible that it is the right way to account for *deductive* applications of epistemic rationality. For on the face of it the deductive case is merely a special case of a much more general question about epistemic rationality. The same point applies in the case of instrumental rationality. If Ronnie is keen on dancing, then the fact that there will be dancing at the party can give him a reason to go there, even if going there isn't *necessary* for him to dance. The Wide-scope accounts of objective and subjective instrumental rationality, like the Wide-scope account of epistemic rationality, are designed to fit the extreme case, in which an action really is or is believed to be *necessary* for an end. But instrumental rationality has a much broader scope. It can be affected in many ways by relationships between actions and ends that fall far short of necessity. If Wide-scope accounts aren't well-suited to explain what goes on in the less strict cases, then they can't be the right account of the strict cases. For the strict cases are surely simply a limiting case of the less strict ones.

7.5.2 Agent-Neutrality

Phil believes that p, that if p then q, and is wondering whether q. A Narrow-scoper can say what reason Phil has to believe that q: after all, he believes that p and that if p then q. This is what gives him reason to believe that q, because that p and that if p then q are the right kinds of thing, if true, to be evidence that q. But according to the Wide-scope view, evidence and reasons for belief have nothing to do with epistemic rationality. Epistemic rationality involves complying with a single, eternal, agent-neutral requirement not to have inconsistent beliefs. Insofar as he is complying with the Wide-scope requirement, when Phil forms the belief that q in this situation, the reason for which he believes q is not that p. Nor is it that p and that if p then q. It is that epistemic rationality requires not believing that p and that if p then q and wondering whether q and bearing these beliefs in mind and not believing that q. So if the Wide-scope account is right, then agents shouldn't believe for the reasons that are their evidence. They should believe for some non-evidential reason that is eternal and agent-neutral and unexplained.

7.5.3 Prima Facie *Reasons*

As in each other case, I think that so long as we are in the business of offering *alternatives* to the Narrow O principles, rather than in looking for disambiguations of them, there is a viable—and perhaps preferable—alternative to Wide-scoping. Instead of weakening Narrow Bf(sg)O by widening the scope of the "ought", we can weaken it by replacing the "ought" with a weaker normative concept.

As suggested by the last subsection, I hold that we should think of the account of epistemic rationality on a par with the account of *subjective* rationality. As in the practical

case, it looks like we can draw a parallel between *objective* and *subjective* senses of the words "reason" and "evidence". If Carrie is 99% reliable, then the fact that Carrie said that *p* is evidence that *p*. It is a reason to believe that *p*. But if we don't know that Carrie has said that *p*, then this isn't evidence that we have. It isn't among our reasons. On the other hand, if we do believe that Carrie said that *p*, then we do have evidence that *p*—we have some reason to believe that *p*. And we can have this reason, even if we turn out to actually be mistaken about whether Carrie said that *p*. Whether Carrie actually said that *p* makes no difference as to how rational it is for us to conclude that *p* on the basis of our *belief* that she did.

If this is right, then the connection between beliefs and epistemic rationality should be much like the connection between beliefs and instrumental rationality. It should be[25] that believing something that is evidence that *p*, if it is true, is how to have a *subjective* epistemic reason to believe that *p*. And then we should say that epistemic rationality just has to do with what you have subjective epistemic reason to believe:

Narrow BfgenSR: If you believe that *p* and *p* is, if true, evidence that *q*, then you have a subjective epistemic reason to believe that *q*.

This *general* account yields as a trivial consequence the account that we were looking for:

Narrow Bf → SR: If you believe that *p* and that if *p* then *q*, then you have a subjective epistemic reason to believe that *q*.

Unfortunately, however, in the epistemic case we clearly need to be yet a little bit more careful.

The problem is this. You may believe something that, if true, is evidence that *q*. But you may also believe something that would *defeat* that evidence. For example, you may believe that Carrie has just said that *p*. But you may also have it on good authority—from Carrie herself, say—that she was simply joking. Though by itself, the fact that Carrie just said that *p* is the right kind of thing to be evidence that *p*, you believe something that cancels the force of this reason. Intuitively, we would say that you have no reason at all to believe that *p*, in this situation.

So if we want to keep a Narrow-scope account, we need to weaken our principle yet further. We should distinguish between *prima facie* and *pro tanto* reasons. You have a *pro tanto* reason to believe that *q* just in case you ought to believe that *q*, if you have no countervailing evidence. Likewise, you have a *prima facie* reason to believe that *q* just in case you have a *pro tanto* reason to believe that *q*, so long as your reason is not completely undermined. A *pro tanto* reason can be partially undermined, but if it is *completely* undermined, then we say that it is merely a *prima facie* reason.

[25] More or less. Obviously we need complications to explain how background knowledge can affect what counts as evidence for you, and so on.

The distinction between partial undermining and complete undermining is important. On a standard example, you see Tom Grabit come out of the library, pull a book from under his shirt, and scurry off. This is reason for you to believe that Tom just stole a book from the library. But if Tom has a twin brother Tim from whom you could not distinguish him, that can undermine your reason to believe that Tom just stole a book. For all you know, it could have been Tim. Yet your reason is not completely undermined in this case. This is easy to see, by observing that it can still get *worse*. For Tom and Tim might have a third identical sibling, Tam, whom you can visually distinguish from neither. If so, then your reason would be still worse. So it can't have been completely undermined. This contrasts with what happens if you look to the side, and notice that a movie director has his camera focused on Tom, and is just saying "Cut! One more take, Tom." If you see that, then it completely undermines your visual evidence that Tom stole a book. Such a reason would be merely *prima facie*.

The plausible Narrow-scope account of epistemic rationality replaces Narrow BfgenSR with Narrow BfgenPF—a principle exactly like it, but with the modifier that the reasons that it invokes are merely *prima facie*.

7.6 The Scope of Instrumental Reason

In each domain in which Wide-scoping is applied, the issues are somewhat different. Plausibly, in each domain the debate between Wide-scopers and Narrow-scopers should be adjudicated on its own grounds. And in every domain, as I've noted at a few points in footnotes, it is possible to introduce further, mixed, theories, with some Wide-scope and some Narrow-scope elements. This complicates issues in ways that I haven't had opportunity, here, to explore. Yet some issues have served as common threads to each domain of inquiry. The first is that the Narrow principles allow "detaching"—if we take as a premise the antecedent of their conditionals, we can simply apply *modus ponens* and deduce results that are largely unacceptable, at least if we restrict our attention to a relatively *strict* normative concept like that expressed by "ought". This is the original motivation for Wide-scoping.

Another issue that raised its head in every domain was the symmetry predictions of the Wide-scope accounts. Every Wide-scope principle yielded predictions of symmetry, and many of these failed to be substantiated. Some of these failures were worse than others. Some had a bearing on whether Wide-scoping can plausibly be generalized to less strict cases, some a bearing on dynamic questions about rationality over time, and some a bearing on simple distinctions applicable at a time, like that between whether an obligation has been *satisfied*, or merely *escaped*.

The third issue that returned over and again was the commitment of Wide-scope accounts to unexplained, eternal, agent-neural requirements that can be avoided by Narrow-scope accounts. Again, the issues were different in different domains. Such requirements are *controversial*, because on one common way of thinking about the

relationship between agent-relativity and agent-neutrality, agent-neutral requirements need to be explained by agent-relative ones. In the account of objective instrumental rationality, we saw that this feature made Wide-scoping unacceptable to adherents of a supposedly widespread view about reasons—that *all* reasons must be explained by an account of objective instrumental rationality.

But the commitment to these requirements is controversial in its own right. For it involves postulating distinct requirements in order to account for each domain. But I've shown how to subsume the accounts of objective and subjective instrumental rationality under one account, and how to account for epistemic rationality and for the role of conscience merely by offering an account of the relationship between objective and subjective reasons. These accounts didn't work by postulating new and unexplained requirements—they worked simply by trying to *make sense* of a group of normative concepts weaker than that expressed by "ought". I haven't had the space, here, to flesh out the details of any of the Narrow-scope alternatives that I've offered, but I hope that I've set out just enough that they can be seen as viable alternatives which fit into an attractive context.

Since much has recently been made of it, in particular, allow me to return to the Wide-scope account of objective instrumental rationality. I want to explore yet one more way in which Wide-scoping can be and ought to be *controversial*. I take it that Ronnie and Bradley differ with respect to what they have reasons to do. In particular, the fact that there will be dancing at the party tonight is a reason for Ronnie to go there, but not a reason for Bradley to go there. And this difference between Ronnie and Bradley seems to be due to a difference in what they like or desire or care about.

The "Humean" Theory of Reasons, as I indicated, holds that all reasons are explained in the same way as Ronnie's. It is a natural view to hold, if you think that at *some* level, all reasons must be explained in the same way. For once you grant that Ronnie's reason needs to be explained by his desire, then the challenge is on to explain just how this could work, so that a desire is required to explain Ronnie's reason, but that is a non-essential part of the uniform explanation of reasons generally.

But as I indicated in section 7.1, the "Humean" Theory of Reasons runs into trouble in more than one way, when it comes to the Wide-scope account of objective instrumental rationality. The argument from Jean Hampton discussed in section 7.1 is only one such way. Stephen Darwall offers a distinct argument that the "Humean" Theory of Reasons is doomed to incoherence. The argument is this: (1) the correct account of objective instrumental rationality is Wide-scope. (2) From Wide ObjO and "you desire that *p*" it does not follow that you have a reason to do what promotes *p*. (3) To derive this conclusion from Wide ObjO, you need to adduce the premise that you ought to desire that *p*.[26] So (4) there can only be reasons for agents to do

[26] This assumption is controversial, and may be false. Given a suitable version of the principle that "ought" implies "can" and the assumption that it is sometimes practicably impossible to give up a desire, it may turn out that the only way to comply with Wide ObjO is to take the means to such a desire. Or it may be that

particular things, such as go to the party, if there are antecedent reasons for them to have certain ends. Therefore (5) on pain of regress, not all reasons can be explained in this way. And so the "Humean" Theory of Reasons, since it claims that they can, is incoherent.

Darwall doesn't use the fact that Wide ObjO itself posits an unexplained agent-neutral requirement in order to create trouble for the "Humean" theory. He merely uses the fact that it doesn't allow us to "detach" in order to create the trouble. So a view like the "Humean" Theory of Reasons *needs* to make sense of a Narrow-scope theory of objective instrumental rationality. It needs an explanation of Ronnie's reason that will not advert to a further reason, on pain of regress. As indicated, I hold that we can do so, if we remind ourselves of the difference between saying that there is a *reason* for someone to do something, and saying that she *ought* to do it. The former is a far weaker claim.

If Darwall and Broome and others are right, and the Wide-scoping program truly is uncontroversial, then that does it for the "Humean" Theory of Reasons. But this argument of Darwall's raises its own bit of trouble for the Wide-scoping program. As I understand the "Humean" theory, it is motivated by an attempt to explain where merely agent-relative reasons come from. Wide-scoping is part of a general picture that takes agent-neutral reasons for granted, as not requiring any particular kind of explanation. It tells us something about what is going on in Ronnie's case. But if Darwall's argument works, then Wide-scopers can only believe that there is a reason for Ronnie to actually go to the party, if they believe in an antecedent reason for Ronnie to desire to dance.

But now we get to the heart of the matter. Is this reason to desire to dance one that is a reason for Bradley as well? If it is, then we haven't succeeded in explaining why Ronnie has a reason to go to the party that isn't a reason for Bradley to go there. But if it is not a reason for Bradley, then the question about where merely agent-relative reasons come from simply retreats another step. According to the Wide-scoper about objective instrumental rationality, they don't arise as a result of differences in desires. But where *do* they come from? On the face of it, wherever we think merely agent-relative reasons come from, that is going to be a domain in which it looks like how things are for some agent can have an effect on what she has reason to do. In short, it is the kind of domain in which Wide-scopers tend to apply their theories.

If Darwall's argument really works against the "Humean" Theory of Reasons, then it is an equally good argument that Wide-scoping has to stop *somewhere*. Since there really are reasons that are merely agent-relative, they have to arise *somehow*. Some differences between agents have to actually lead to a difference in what people ought or have reasons to do—one that can eventually explain the difference between Ronnie

there is some special requirement not to give up ends, once you have them. These would be ways in which (3) would be false, but they would also be ways in which (2) would be false. See Greenspan [1975] for an excellent discussion of the first kind of possibility.

and Bradley. The "Humean" simply figures that we might as well stop at the beginning. For the sophisticated "Humean", Narrowing the scope of instrumental reasons is just the trick in order to plausibly *expand* the scope of instrumental *reason*. And if very much at all of this paper is correct, then it's not at all obvious that this project is crazy or "confused" or narrow-minded. Whether any Wide-scope account is correct or not should still be very much a live issue.[27]

[27] Special thanks to Jim Pryor, Ant Eagle, Gideon Rosen, David Sussman, John Hawthorne, Stephen Darwall, John Broome, and Stephan Leuenberger.

8

Means–End Coherence, Stringency, and Subjective Reasons

Intentions matter. They have some kind of normative impact on our agency. Something goes wrong when an agent intends some end and fails to carry out the means she believes to be necessary for it, and something goes right when, intending the end, she adopts the means she thinks are required. This has even been claimed to be one of the only uncontroversial truths in ethical theory. But not only is there widespread disagreement about *why* this is so, there is widespread disagreement about in *what sense* it is so. In this paper I explore an underdeveloped answer to the question of in *what sense* it is so, and argue that resolving an apparent difficulty with this view leads to an attractive picture about *why* it is so.

8.1 Means–End Coherence

8.1.1 The Problem

Zach intends to do some action, *A*. And he believes that to do *A*, he must do *B*. Zach bears an interesting and important normative relationship to *B*. It is an action that he believes to facilitate his intended end, and something is going wrong, if he intends *A*, believes *B* to be necessary for *A*, has reflected clear-headedly on this fact, and yet still fails to intend to do *B*. But it turns out to be hard to say exactly what this relationship is, or what, exactly, is going wrong with Zach in this situation.

It is not, for example, that Zach ought to do *B*, as one might hold on a naïve view. For example, suppose that *A* is the action of hiring an assassin to murder his wife, and *B* is the action of paying the assassin. It is not, plausibly, true that Zach ought to pay the assassin. Nor is it plausibly true that Zach ought to intend to pay the assassin. But there is still something wrong with Zach if he intends to hire an assassin to murder his wife, believes that paying the assassin is necessary in order to do this, and has no intention whatsoever of paying the assassin, no matter how long and clear-headedly he thinks about the matter.

Philosophers have constructed a number of fantastic ideas about what is going on in this case, instead. According to *Wide-scopers*, Zach ought to do some *disjunctive*

action—to either-intend-to-pay-the-assassin-or-not-intend-to-hire-him-to-murder-his-wife.[1] It is because Zach does not satisfy this disjunction, that there is something wrong with him.[2] According to others, there is nothing *normative* going astray with Zach, in the narrow sense of 'normative' in which it contrasts with 'evaluative'—rather, there is simply a failure of function of some kind. It is a malfunction of agents to not intend the means they believe to be necessary, but it doesn't reflect on them, personally; rather it is more like having a leaky heart valve.[3] And according to yet another view, it is true that Zach ought to intend *B*, but not in the ordinary, practical, sense of 'ought'. Rather, it is true in the *epistemic* sense of 'ought'. Kieran Setiya motivates this view as the last resort—on the grounds that none of the other ways of describing what is going on in Zach's case even makes sense.[4]

I think, however, that all of these ideas, in addition to having their own problems, pass over a very simple one which possesses great promise. On the view I'll suggest, Zach *subjectively* ought to do *B*, and *subjectively* ought to intend *B*. In the remainder of part 1 I'll motivate this idea's initial promise. Then in part 2 I'll explain an important challenge that it faces—the challenge which Setiya has argued is fatal to the view, although I will state the challenge in a way that does not depend on Setiya's substantive views about reasons. In part 3 I'll explain how we can get around this challenge, provided that we are willing to make a certain kind of assumption about the nature of intention. The resulting view has the virtue, moreover, of having the right structure to also give us an elegant and complete explanation of *why* this relationship holds between Zach and his intentions. Those will be the main positive ideas of the paper. Then in part 4, I'll start to refine these ideas, showing how to make the required assumptions about intention more plausible, and consider the advantages for an important phenomenon about deliberation over time. Finally, in part 5, I'll show how the resulting view is able to accommodate and explain the relationship between intentions and necessary means as a limiting case of the relationship between intentions and merely facilitating means. Because this is a deep problem for other accounts, I'll suggest that the ease with which my account deals with it is collateral evidence that the account is on the right track.

8.1.2 Ewing's Problem and a Working Hypothesis

Our problem has the following structure: something goes wrong under certain circumstances. And those circumstances depend on what the agent believes. It is the actions Zach *believes* to be necessary means to doing *A*, to which he bears the interesting and important normative relationship—not the actions which *are* necessary means to doing *A*. So in looking to understand what is going on in Zach's case, it should

[1] Possibly adding '-or-not-believe-that-paying-the-assassin-is-necessary-to-hire-him', as well. See Chapter 7, this volume, for discussion.

[2] See, for example, Hill [1973], Gensler [1985], Wallace [2001], and Broome [1999], [2001].

[3] See, for example, Bratman [1987], [2009] and Raz [2005a].

[4] Setiya [2007a]. I should note that several of the authors cited in fact hold some combination of these views.

be fruitful to examine other cases in which something goes wrong in circumstances that depend on what the agent believes. And fortunately, there are a number of other important cases in ethical theory with this very structure. Here is one: Raul believes that there is a vast, international conspiracy whose primary objective is to trick him into playing hopscotch. Raul needs professional help. His need for help depends on his beliefs. Saul, who has no such belief, does not have Raul's pressing need for professional help. Something is going wrong, if Raul does not get help, but not wrong, if Saul does not get help.

The thing that goes wrong in Raul's case, however, does not seem to be very much like the thing that goes wrong in Zach's case, if he does not intend *B*. For example, an impartial and benevolent bystander with all relevant information about Raul's case would advise him to seek professional help.[5] But an impartial and benevolent bystander with all relevant information would not advise Zach to intend to pay the assassin. She would advise Zach to give up on his plan to hire an assassin to murder his wife. So in looking for cases to compare Zach's to, we should avoid cases like Raul's—something different seems to be going on, in them.

There are other important cases, however, that seem to differ from Raul's in the same way that Zach's case differs from Raul's. For example, there is an important relationship between an agent and the things she *believes* she ought to do. Suppose that Yves believes that he ought to do *A*, but does not do *A*. Something is going wrong in such a case. And as in Zach's case and Raul's case, it is something that depends on Yves's beliefs. But it does not seem to depend on his beliefs in the same way that Raul's case does. As we observed before, an impartial and benevolent bystander with all relevant information about Raul's case would advise him to seek help. But an impartial and benevolent bystander with all relevant information about Yves's case would not necessarily advise him to do *A*. For example, Yves's belief that he ought to do *A* might be *false*. It might even be that Yves ought *not* to do *A*. In that case, a benevolent and impartial observer would surely advise Yves *not* to do it.

So Yves's case is structurally like Zach's case in many ways. In each case, something can go wrong, which depends on their beliefs, but it is not the sort of thing that goes wrong in Raul's case, or else does not depend on their beliefs in the same way. Moreover, in neither case is it true that they ought to do the thing such that, when they do not do it, something is going wrong. This is an important point. It sounds platitudinous to say that you ought to do what you believe you ought to do—that's just to say that you ought to let your conscience be your guide. But it can't be in general true that for each thing anyone believes she ought to do, she ought to do it. For if it were, then everyone would be infallible about what she ought to do. Which is false. People are not, in general, infallible about what they ought to do. But there is still some interesting relationship between them and the things that they believe that they ought to do.

[5] Compare Schroeder [2007a].

Saying exactly what this relationship is, is known as *Ewing's Problem*.[6] A number of philosophers have drawn attention to its parallel structure to the problem about intentions with which we began.[7] According to a Working Hypothesis that it would be worthwhile to investigate, the same sort of thing is going on in each case. That is the Working Hypothesis that I am going to investigate in this paper. I will be defending the view that the Working Hypothesis is correct. In the next two sections, I will defend a version of Ewing's answer to his own problem, in contrast with the competing Wide-scope answer. Then in the remainder of the paper I will show how this answer can be extended to the case of means–end coherence, in a way that yields not just an answer to *in what sense* something is going wrong with Zach, but an answer to *why*.

8.1.3 The Wide-scope Answer

Wide-scopers like Broome [1999] have also been keen on the parallel between the problem about intentions and Ewing's Problem. They claim that their strategy works for both. But it is a miserable answer to Ewing's Problem. The Wide-scope view is that what Yves ought to do, is to either-do-*A*-or-not-believe-that-he-ought-to-do-*A*. This is why, they say, something is going wrong with him if he believes that he ought to do *A* and doesn't do *A*. He is falling astray of this particular thing that he ought to do—the one that says that he shouldn't be that way.

The advantage of the Wide-scope answer to Ewing's Problem is that it is supposed to explain why it does not follow that Yves ought to do *A*. All that he ought to do is to either-do-*A*-or-not-believe-that-he-ought. And from this, they claim, it does not follow that he ought to do *A*. Another perfectly good way of doing this disjunctive thing is to change his belief. As a result, Yves's belief that he ought to do *A* is not, after all, infallible, and hence we have a solution to Ewing's Problem. I think there are at least three very bad problems for this Wide-scope answer. The first is that it is symmetric in a way that the duty of conscience is not. And the second two are both ways in which Wide-scopers *are*, after all, committed to concluding that in at least some kinds of case, Yves's belief really is infallible—the very problem Wide-scoping is motivated in order to avoid.

The first problem for Wide-scoping is that it is symmetric. It doesn't distinguish between acting in accordance with your moral beliefs and adopting moral beliefs in accordance with your actions, and as a result it fails to distinguish between following your conscience and the distinctive vice of rationalization.[8] Rationalization is the vice of changing your beliefs about what you ought to do, because you are not going to do it, anyway. According to the Wide-scope view, this is precisely as good a way of satisfying

[6] Ewing [1953], Piller [2007].

[7] For example, Gensler [1985] and Broome [1999]. I discussed both kinds of case in Chapter 7, this volume, although I now think that I was wrong to say that my remarks about 'desire' there apply equally well to the case of intentions.

[8] See Chapter 7, this volume, for the statement of this objection. Kolodny [2005], [2007] also presses a version of this kind of objection against a kind of Wide-scope view.

this requirement as is actually paying attention to what you believe and acting accordingly. If the thing that is going right about Yves when he believes that he ought to do *A* and as a result, does it, is also going on the case of the Ultra-Rationalizer, who never lets her actions be affected by her beliefs about what she ought to do, but rather immediately changes her mind about whether she ought to do them, then it seems to me that we have not quite captured what is going wrong in Yves's case.

A naïve response to this objection holds that though changing his beliefs is not ruled out by the instrumental principle, it *is* ruled out by some *other* principle governing *theoretical* reason, which says not to change your beliefs about what you ought to do, or some such thing. But this is short-sighted. In general, if Yves ought to do either *A* or *B*, and ought not to do *B*, then it follows that Yves ought to do *A*. This means that if there really is a principle according to which Yves ought not to change his belief, then in any circumstance in which that principle applies, the only way for him to fulfill his Wide-scope requirement to either-do-*A*-or-not-believe-that-he-ought is for him to do *A*. And so it follows that in those cases, he ought to do *A*. But that just means that in any case in which Yves ought not to change his belief, it follows from the Wide-scope view that his belief must be true. So the Wide-scope view is still committed to holding that Yves's belief about what he ought to do is infallible in any case in which he is believing rationally. And this looks like a second bad problem for the Wide-scope view.

Moreover, it follows from another plausible general principle that things are even worse.[9] As Patricia Greenspan has pointed out, it follows from the fact that Yves ought to either do *A* or do *B* and that Yves *cannot* do *B*, that Yves ought to do *A*.[10] But plausibly, in at least many cases, changing his beliefs about what he ought to do is not something that Yves can do as instantaneously as he can act on them. Even if he has the power to change those beliefs, doing so takes some time. Consider, then, the intervening time. That is time during which the only way that Yves can either do *A* or not believe that he ought, is to do *A*. So it follows from the Wide-scope view and this principle that during that time, he ought to do *A*. But with respect to all of our beliefs that we are not already trying to change, we are *always* in this intervening time. Hence, it follows that people are in general infallible about what they ought to do, as long as they do not try to change their minds. Whenever they believe they ought to do something and are not already trying to change their mind about it, what they believe is true—they really ought to do it. Again, to accept this conclusion is to have failed to solve Ewing's Problem in the first place.

These, I think, are each very bad problems for Wide-scoping as an answer to Ewing's Problem. So if the hypothesis that the problem about intentions and Ewing's Problem should receive similar answers is a good one, that means that Wide-scoping can't be right about intentions, either. (And indeed, the Wide-scoping account of the case of

[9] Setiya [2007a] also advances a version of this argument, although mine requires weaker assumptions, because it takes into account time. Full credit for the argument goes to Greenspan [1975].

[10] Greenspan [1975].

intentions faces problems in its own right.[11]) It is Ewing's own answer to his Problem that gives us a more promising way of developing our Working Hypothesis.

8.1.4 Subjective Oughts

Ewing's own answer to his Problem is that 'ought' is ambiguous in English, having two important senses. It is true, Ewing held, that you ought to do what you believe you ought to do. But this does not mean you are infallible about what you ought to do, because the first 'ought' has a different meaning from the second. There is an *objective* sense of 'ought', and you have beliefs about what you ought to do in that sense. But there is also a subjective sense of 'ought', and it is true that you subjectively ought to do those things that you believe that you objectively ought to do.

Ewing also had a simple explanation of why this last thesis is true. He held that the subjective sense of 'ought' just *meant* 'believes she objectively ought', making the requirement to follow your conscience simply definitional of the subjective sense of 'ought'. We needn't go in for this idea of Ewing's about how to understand the relationship between the two senses of 'ought', however, in order to see the attractions of his solution.

Moreover, it is a virtue of this solution that Ewing's Problem is only one kind of case which warrants distinguishing a subjective sense of 'ought'. In Ewing's cases, someone has a belief directly about what she ought to do, but which is mistaken. But in many cases, agents have no explicit beliefs about what they ought to do. Still, even in those cases, there are things that they ought to do. And the things that they ought to do sometimes depend on features of their situation. In some situations we ought to do some things, while in other situations we ought to do others. And finally, the very features of our situations on which what we ought to do depend, can sometimes be features of which we are unaware, or even have false beliefs about.

Bernard Williams' gin and tonic case is such a case.[12] We are to imagine that Bernie is in the sort of situation in which the thing for him to do, if his glass contains gin and tonic, is to take a sip. If his glass contains gasoline, on the other hand, he ought not to take a sip. So what he ought to do depends on whether his glass contains gin and tonic or gasoline. We are also to imagine that Bernie believes that his glass contains gin and tonic, but that in reality, it contains gasoline. Philosophers disagree about what Bernie ought to do. According to some, he ought to take a sip, for it is the only rational thing for him to do—the best option, given his beliefs. According to others, he ought not to take a sip, since gasoline is toxic. But very plausibly, according to a generalization of Ewing's solution, these philosophers are simply cottoning on to different ways in which we can use the word 'ought'. Sometimes we can use it in the objective sense, in

[11] I do think that the problems I've pursued are *more pressing* in the case of Ewing's Problem, which is part of why I think it generates leverage to tackle the problem indirectly, by way of our Working Hypothesis.

[12] Williams [1981]. See Schroeder [2008] for extensive discussion.

which he ought not to take a sip. But it also has a subjective sense, in which he ought to take a sip. Both sides are right—about something. For there are two senses of 'ought'.

In truth, for our purposes it doesn't matter greatly whether there are really two senses of 'ought'. It doesn't really matter whether or not it is really okay to use the word 'ought' for both of these relations—though if it were, that would yield the best explanation of why 'you ought to follow your conscience' makes perfect sense. What matters to me is simply to point out that there is an interesting normative relationship of some kind between Bernie and taking a sip, and another interesting normative relationship between Bernie and not taking a sip. Both relationships are interesting, and we appeal to them for different purposes. For example, our practices of advice hinge much more heavily on the latter relationship, and our practices of blame and praise depend much more heavily on the former. I'm going to call them the subjective ought and objective ought relations, but that is just to give them names. If you think one or the other doesn't deserve to be called 'ought', I'm happy to defer, and you may call it something else.

The important point is that once we get past the obstacle of thinking that the interesting claim here is semantic, I think the claim that there is a subjective 'ought' leaves little about which to be alarmed. Everyone should agree that there is an interesting normative relationship between Bernie and taking a sip. My hypothesis about Ewing's Problem is that the same relationship holds between someone who believes that she ought to do something, and what she believes that she ought to do. This is the 'in what sense' answer to Ewing's Problem.

8.2 A First Pass

8.2.1 The Explanation of Ewing's Problem

So far, I've isolated an answer to *in what sense* something is going wrong in Yves's case, but I haven't explained *why*. Ewing had a simple explanation of why. According to Ewing, 'Yves subjectively ought to do *A*' just *means* 'Yves believes that he objectively ought to do *A*'. That claim about what 'subjectively ought' means makes it trivial that you subjectively ought to do what you believe you objectively ought to do. But it doesn't accommodate Bernie's case. Bernie doesn't have any explicit beliefs about what he ought to do, but there is still something that he subjectively ought to do.

Reflection on Bernie's case, however, motivates a more general principle about the relationship between objective and subjective oughts, that provides an alternative to Ewing's explanation. What is distinctive of Bernie's case, after all, is not that he believes that he objectively ought to do something, but that he believes something which, if true, entails that he objectively ought to do it. In general, the following seems[13] like a plausible principle:

[13] In section 8.5 I'll consider and endorse an argument that this principle can't be exactly right, and show how to introduce refinements which get around this problem, but for now it is close enough. Certainly it has

subjective ought test *X* subjectively ought to do *A* just in case *X* has some beliefs which have the following property: the truth of their contents is the kind of thing to make it the case that *X* objectively ought to do *A*.[14]

The subjective ought test is a general principle about the relationship between objective and subjective oughts. It is motivated by considering a range of cases that are more general than the pure Ewing's Problem cases like Yves's—including Bernie's. But it still provides us with an explanation of why Yves subjectively ought to do *A*.

This is because Yves's case passes the subjective ought test perfectly. In Yves's case, what he believes is that he objectively ought to do *A*. But if that is true, then it *is* the case that he objectively ought to do *A*. So if the subjective ought test is true, it is a general principle which can explain why Yves subjectively ought to do *A*:

P1 Yves believes that he objectively ought to do *A*.
P2 If this belief is true, then Yves objectively ought to do *A*.
P3 If Yves has some beliefs which, if true, entail that he objectively ought to do *A*, then Yves subjectively ought to do *A*.
C Yves subjectively ought to do *A*.

The explanation is simple. Premise 1 simply describes Yves's case: he believes he ought to do something. Premise 2 is trivial. And Premise 3 is the right-to-left direction of the subjective ought test. So if the right-to-left direction of the subjective ought test is true, then Yves subjectively ought to do *A*. I take this to be an explanation, because it subsumes Yves's case under a much more general principle—one which explains what is going on in Bernie's case, as well.

So where does that leave us? Recall that our Working Hypothesis was that what is going on in Zach's case is much like what is going on in Yves's case. And I've just shown how what is going on in Yves's case can be explained by appeal to general principles which apply to an even wider range of cases. If our Working Hypothesis is right, then we should conclude that Zach subjectively ought to do *A*, and if we want to understand *why*, we should look, as in Yves's case, to our general principles about how subjective oughts work. But this leads to a problem.

8.2.2 The Hitch

We can explain Yves's case by appeal to the subjective ought test, because the belief on which what Yves subjectively ought to do depends, is such that if it is true, then Yves objectively ought to do *A*. The hitch is that Zach's case is not like that. The belief on which Zach's situation depends is his means–end belief: that to do *A* he must do *B*. But

been widely accepted—for example, by Parfit [2011], who calls the distinction between objective and subjective oughts the difference between reasons and rationality.

[14] Sometimes tests like this one are formulated in counterfactual terms. It is easy to see, though, that the counterfactual test can't really be quite right. For example, it could be that the closest world in which Bernie's glass really contains gin and tonic is one in which he has promised to give up on drink, or satisfies some other condition which would make a difference to what he objectively ought to do.

that belief can be true, even though it is not the case that Zach objectively ought to do *B*. For it may simply not be the case that he objectively ought to do *A*. To return to our example, it is not too hard to imagine that Zach's belief is true, that to hire an assassin to murder his wife, he must pay the assassin. But it is still clearly not the case that Zach objectively ought to pay the assassin.

Moreover, this doesn't merely present an obstacle to our *explaining* why it is that Zach subjectively ought to do *B*. It presents an obstacle to its even being *true*. For according to the left-to-right direction of the subjective ought test, Zach subjectively ought to do *B only if* he has some beliefs which, if true, guarantee that he objectively ought to do *B*. But all that we know about Zach is that he intends to do *A* and believes that to do *A* he must do *B*. And it is easy to imagine cases in which this belief is true, but it is not the case that Zach objectively ought to do *B*.

In fact, we can raise the same problem by appeal to a principle that is much weaker than the subjective ought test. Even if the subjective ought test turns out to be false, the following principle is still highly compelling:

very weak ought test: If *X* is completely opinionated about every factual question, has a complete credence (=1) in every proposition she believes, and all of her beliefs are true, then if *X* subjectively ought to do *A*, it follows that she objectively ought to do *A*.[15]

Now imagine that Zach is completely opinionated about every factual question, has a complete credence in every one of those beliefs, and that all of his factual beliefs are true. And imagine that he intends to hire an assassin to murder his wife and believes, truly, that paying the assassin is necessary to hire her. Then from the supposition that Zach subjectively ought to intend to pay the assassin, it would follow that Zach objectively ought to pay the assassin. But that seems false. Even in this case, there is something going wrong with Zach if he does not intend the means. But not even in this case is it plausible that Zach objectively ought to intend the means.

The problem is that our examples suggest that there is some close connection between subjective and objective oughts which is mediated by the beliefs that the subjective oughts depend on. And in the case of means–end coherence, Zach's situation seems to depend on his means–end belief. But there does not appear to be a connection between subjective and objective oughts in Zach's situation that is mediated by this belief. So it is hard to see how it could be true that Zach subjectively ought to intend the means.

This, in essentials, is Kieran Setiya's argument against the idea that Zach subjectively ought to intend to pay the assassin, in his forthcoming paper, 'Cognitivism About Instrumental Reason'.[16] Setiya claims that this view is no better than the view that Zach objectively ought to intend to pay the assassin, because it entails it, anyway. Which is just what we've seen follows, given that Zach's belief is true, from the subjective ought

[15] The qualifications in the very weak ought test address a class of proposed counterexamples to the subjective ought test that I will discuss in part 5.

[16] Setiya [2007a].

test. Setiya doesn't use the subjective ought test; he appeals to his own substantive analysis of reasons in terms of good practical reasoning.[17] But I think that the problem I've outlined here is just a generalization on his argument in that paper, appealing to weaker assumptions.

8.2.3 How Bad

The problem, I think, is not a trivial one for our hypothesis. It tells us that even the view about subjective 'oughts' is committed to the objectionable conclusion about what Zach objectively ought to do that we initially sought to avoid, at least in the special case in which Zach's means–end belief is true. I have elsewhere advocated a very general strategy for dealing with such unintuitive results,[18] so I want to say briefly why I think that strategy would not work, here.

The strategy I have employed elsewhere involves weakening the claim that Zach objectively *ought* to intend to pay the assassin, to the weaker claim that there is an objective *reason* for Zach to intend to pay the assassin.[19] The next step of the strategy is to distinguish between this bare existential claim, and the further thesis that it is a particularly weighty reason.[20] The strategy is to deny, in the cases of unintuitive reasons, that they are very weighty at all, insisting that they are of particularly low weight, and trying to explain why.[21] And finally, the strategy is to explain why it seems like there is no reason at all for Zach to intend to pay the assassin, by explaining why pragmatic factors dictate that our negative existential intuitions about reasons are likely to be systematically misleading, in the case of reasons of very low weight, and to provide independent evidence that this is so.[22]

Although I have never committed in print to any views specifically about the norms governing intentions, for some time I did wonder whether something like this strategy could be an important part of solving the problem of what is going on in Zach's case. There are a variety of reasons why I think this can't be right. But the principal one, is that given a generalization of our test for subjective 'oughts', it fails miserably to account for the *stringency* of what goes wrong with Zach when he fails to intend the means he believes to be necessary to his intended end.

The distinction between objective and subjective reasons can be made in the same ways as that between objective and subjective 'oughts'. It can be done by considering Ewing-style cases, or by focusing on Bernie's case.[23] Because they are motivated by similar cases, similar tests seem to apply:

subjective reason test: *X* has a subjective reason to do *A* just in case she has some beliefs which have the property, if they are true, of making it the case that *X* has an objective reason to do *A*.

[17] See Setiya [2007b].
[18] Chapter 7, this volume, Schroeder [2005a], [2007a], [2007c].
[19] Chapter 7, this volume.
[20] Schroeder [2005a], [2007c], and especially [2007a, chapter 5].
[21] Schroeder [2007c] and especially [2007a, chapter 7].
[22] Schroeder [2005a, 6–11], [2007a, chapter 5], [2007c].
[23] See Schroeder [2008].

Moreover, it seems plausible to suppose, in connection with the subjective reason test, that if the objective reason would be weightier, then the subjective reason is weightier. At any rate, I'll work with this idea.

Here is the problem: the interesting normative relationship between Zach and intending to pay the assassin is a *stringent* one, a *requiring* one. It is very strong. Something goes *very* wrong with Zach, if he clear-headedly intends to hire an assassin, recognizes that hiring an assassin requires paying her, and has no intention whatsoever to pay her. So by our test, something should be going *very* wrong in the objective sort of way with Zach, if he intends to hire an assassin and hiring an assassin really does require paying her, and he has no intention of paying her. So even if we could accept that there is something just a little bit wrong with Zach in that case, because there is *some* objective reason for him to intend to pay the assassin, but only a very weak one—so weak that our intuitions about it might be misleading—that wouldn't be enough. The test requires that there must be something objectively *stringent* going wrong with Zach in the case in which his belief is true, in order for there to be something subjectively stringent going wrong with him, when he has that belief.

The stringency of the connection between Zach and intending to pay the assassin magnifies, I think, the problem posed in the last section. By the subjective ought test and the assumptions that we are making so far, it seems not only to follow from the view that Zach's case is like Yves's and Bernie's cases in that there is some objective reason for Zach to intend to pay the assassin, if his belief is true. It also seems to follow that in order to account for the stringency of the requirement for means–end coherence, this view would be committed to holding that Zach is objectively *required* to intend to pay the assassin. And that is something that we can all agree is patently false. It is not the kind of thing that the strategy I mentioned is capable of explaining away. So we need, I think, another solution.

8.3 A Positive View

8.3.1 Oughts and Transmission

To see how to solve this problem, I think we need to step back and observe a very general fact about the transmission of oughts, which we can use to construct a generalization of Yves's case. This generalization will allow us to see which assumption that we were implicitly making was leading to all of the fuss, and consequently which assumption we can make about intentions that will remove the problem. The fact about the transmission of oughts is simple:

ought transmission: If *X* objectively ought to do *A*, and to do *A X* must do *B*, it follows that *X* objectively ought to do *B*.

In part 5 I'll generalize this principle to reasons, but to get the basic structure of the idea, that is all that we need for now. The idea behind this principle is that the force of

the first 'ought' *transmits* to the second, by means of the necessary connection between them.[24] I'm going to assume that this principle is true for objective oughts.

If it is, then it follows from the subjective ought test that something similar is true for subjective oughts. Suppose that Xera subjectively ought to do *A* and *believes* that to do *A* she must do *B*. Since she subjectively ought to do *A*, by the left-to-right direction of the subjective ought test there must be some things she believes, such that their truth would guarantee that she objectively ought to do *A*. Call the set of those beliefs *S*. But Xera also believes that to do *A* she must do *B*. So consider this belief, along with those in *S*. If all of *those* beliefs are true, then the transmission principle, above, guarantees that Xera objectively ought to do *B*. So since Xera believes these things, it follows from the right-to-left direction of the subjective ought test that she subjectively ought to do *B*. So this induces a *reflection* of the transmission principle for objective 'oughts' for the case of subjective 'oughts':

transmission reflection: If *X* subjectively ought to do *A*, and believes that to do *A* she must do *B*, it follows that *X* subjectively ought to do *B*.

So now consider Xera's case more carefully. Let me stipulate that the content of *S* is her belief that she objectively ought to do *A*. This makes her case a generalized version of Yves's case. Yves's case does not depend on any other beliefs that he has. He simply has the belief that he ought to do *A*, and as a result, subjectively ought to do *A*. But Xera's case is more general. She has a belief about what she ought to do, and she subjectively ought to do this, just like Yves. But due to the reflection of the transmission principle, she also subjectively ought to do other things, which she believes are necessary means to *A*.

But Xera also has an interesting property: even if her belief that *B* is necessary for *A* is true, that is not enough to guarantee that she objectively ought to do *B*. For it could be that her belief that she ought to do *A* is false. So Xera fails the same test as Zach did. She definitely subjectively ought to do *B*—that follows from transmission reflection. But given the transmission principle, this depends on more than one of her beliefs. In order to test it, we must look at *all* of her beliefs that it depends on, not just the transmitting beliefs. It depends both on her belief that *B* is necessary for *A and* on her belief that she ought to do *A*. And if *both* of those beliefs are true, then the transmission principle does guarantee that she objectively ought to do *B*.

8.3.2 A Motivated Conjecture

Xera's case leads to a motivated conjecture: that the same thing might be going on in Zach's case. We had good cause to hypothesize, after all, that Zach subjectively ought to intend to pay the assassin. His case had a lot in common with those of Yves and Bernie, because the thing that was going wrong with him when he didn't so intend seemed to

[24] Compare Darwall [1983, 16], Raz [2005a, 3–9].

depend on his beliefs in a very similar way—quite different from how the thing going wrong with Raul depended on his beliefs. Yet Zach's case still differed from those of Yves and Bernie—it differed from them in exactly the same way that Xera's case differs from them. So it is more than natural to wonder whether Zach's case isn't to be explained in the same sort of way as Xera's is.

If so, then we could explain exactly where Setiya's reasoning went wrong. It went wrong in supposing that Zach's means–end belief is the *only* belief that his subjective 'ought' depends on. This assumption leads to the wrong conclusion about Xera's case, so if Zach's case is like hers in the right way, then it would lead us astray in Zach's case in the same way. In Xera's case, of course, the relevant difference is that her subjective 'ought' depends on her means–end belief because it plays the role of *transmitting* a further subjective 'ought'. So if Zach's case is like that, then his means–end belief must also simply be transmitting a further subjective 'ought'.

At a first pass, then, this leads to the idea of explaining Zach's case in this way:

P1 Zach intends to do *A*.
P2 If Zach intends to do *A*, then he subjectively ought to do *A*.
P3 Zach believes that to do *A*, he must do *B*.
P4 Transmission Reflection: if Zach subjectively ought to do *A* and believes that to do *A* he must do *B*, then he subjectively ought to do *B*.
C Zach subjectively ought to do *B*.

Premises 1 and 3 simply describe Zach's case. Premise 4 is the transmission principle for subjective oughts, which is both independently plausible and derivable (as I showed in section 8.3.1) from the transmission principle for objective oughts and the subjective ought test. And the conclusion follows validly from the premises. So the only premise in need of further defense is premise 2.

So far, I've offered no explanation or defense of premise 2, so this doesn't yet amount to an explanation of why Zach subjectively ought to do *B*—only a model for one. But I hope to have argued that it is a *promising* model, and hence that premise 2 is worth trying to explain. The explanatory model that I am advocating answers the question of *in what sense* there is something wrong going on in Zach's case, if he intends to do *A*, believes that doing *B* is necessary for *A*, and fails to do *B*. This sense is that he is not doing what he subjectively ought to do. Moreover, the explanation on offer answers Setiya's concern—it does not turn out, on this view, that Zach objectively ought to do *B*, even if his means–end belief is true. And finally, it potentially leads to an attractive explanation of *why* Zach subjectively ought to do *B*, which appeals, in addition to whatever assumptions we need in order to explain premise 2, only to plausible, general, independently motivatable principles about *oughts*: the subjective ought test and the objective ought transmission principle.[25]

[25] Alternatively, since the only need we had for the left-to-right direction of the subjective ought test was in order to derive the subjective reflection of the transmission principle from ought transmission, we could make do only with its right-to-left direction, transmission reflection, and premise 2.

So I take it that this kind of explanation would be well-motivated and have several nice features, if only we could explain premise 2 in a principled way. In the next section, I'll consider how to do this. Then in part 4 I'll refine the view, and illustrate one of its important advantages. The view developed in parts 1–4 is, however, only roughly correct. In part 5 I'll explain why the assumptions I'm making are only approximately correct, show how to replace them with more defensible assumptions about subjective *reasons*, and show how the resulting view generalizes very neatly to deal with cases of means–end reasoning to non-necessary means—a serious obstacle to many of the existing views in the literature.

8.3.3 Explaining Premise 2

According to premise 2—the missing step in our explanation—if Zach intends to do something, then he subjectively ought to do it. But is this true? And if so, what would it take to explain it? Fortunately, the subjective ought test gives us the answer. According to the left-to-right direction of the subjective ought test, Zach subjectively ought to do *A* only if he has some beliefs which are such that if they are true then Zach objectively ought to do *A*. So if premise 2 is true, then the following thesis about intention *must* also be true:

nature of intention: If you intend to do *A*, then you have some beliefs which are such that, if they are true, then you objectively ought to do *A*.

So that tells us what *has* to be true, in order for premise 2 to be true. Moreover, nature of intention is also *sufficient* to explain the truth of premise 2. For premise 2 follows from it, together with the right-to-left direction of the subjective ought test.

So is nature of intention true? The simplest thing that you could believe, in intending to do *A*, that would make nature of intention true, is simply that you ought to do *A*. This, of course, is an old and familiar idea about intentions: that they involve or presuppose judgments about what you ought to do. If this old and familiar idea is true, then that would complete our explanation, by making Zach's case a special case of Xera's:

P1 Zach intends to do *A*.
  P2.1 If Zach intends to do *A*, then Zach believes that he ought to do *A*.
  P2.2 If Zach's belief that he ought to do *A* is true, then he ought to do *A*.
  P2.3 Right-to-left direction of the subjective ought test: if Zach has some beliefs such that, if they are true, then he ought to do *A*, then Zach subjectively ought to do *A*.
P2 If Zach intends to do *A*, then he subjectively ought to do *A*.
P3 Zach believes that to do *A*, he must do *B*.
P4 Transmission Reflection: if Zach subjectively ought to do *A* and believes that to do *A* he must do *B*, then he subjectively ought to do *B*.
C Zach subjectively ought to do *B*.

In this explanation, premises 1 and 3 are simply the features of Zach's situation—in which means–end coherence is supposed to apply. Premise 4 is just the transmission

principle for subjective oughts, which is both independently plausible and follows from the transmission principle for objective oughts and the subjective ought test. And premise 2 follows from premises 2.1–2.3, of which 2.2 is trivial and 2.3 is the right-to-left direction of the subjective ought test. So in addition to the transmission principle for objective oughts and the subjective ought test, this explanation appeals to only one substantive assumption: premise 2.1, that intending entails believing that you ought. So if this old and familiar assumption is true, then it yields a powerful explanation of what is going on in Zach's case, which otherwise appeals only to very general principles that can be motivated on the basis of independent cases.

8.4 Defending the Idea

8.4.1 The Problem of Picking

So far I hope to have argued that the assumption that intending involves believing you ought is a *fruitful* one, in the sense that given the kinds of very general background principles given by the subjective ought test and the transmission principle for objective oughts, it is sufficient to explain why Zach subjectively ought to take the means. There are at least three major problems, however, for the view that intending to do *A* entails believing that you objectively ought to do *A*. In increasing order of difficulty, they are the problem of *picking*, the problem of *akrasia*, and what my colleague Jake Ross calls the *three-envelope* problem. In this section I'll explain the problem of picking and argue that solving it requires no real modification in our view. In the next, I'll explain the problem of *akrasia*, and suggest a few different kinds of modification that would allow us to preserve the outlines of this kind of account of the norm of means–end coherence. I'll put the three-envelope problem off until part 5 and consider a couple of loose ends first, as confronting the three-envelope problem will require revisions to the subjective ought test, as well as to the idea that intending involves believing you ought.

The problem of picking is simple. When Zach goes to buy a carton of milk, there are two for him to choose from, the one on the left, and the one on the right. Neither has any advantage over the other, and since he knows this, he knows full well that it is not the case that he ought to take the one on the left, as opposed to the one on the right. Nevertheless, he has to pick one, and so he opts for the one on the left, *without* believing that he ought to have.

On the face of it, this is an intention. Zach has formed an intention to take the carton on the left. But in truth, it doesn't matter whether we call it an 'intention' or not. What matters, is that it is subject to the same norms of means–end coherence as the other cases we are trying to investigate. Once Zach opts for the carton on the left in this way, he must, as he recognizes, open the door on the left in order to get it. So something is going wrong with him if he has no intention whatsoever of opening that door! So it actually doesn't matter whether we call this case an intention or not; we won't have

solved our problem unless our solution applies in this case. And in this case, it seems that Zach does *not* believe that he objectively ought to take the carton on the left.

I think that what cases of picking really show is that the intention to do *A* needn't require the *antecedent* belief that one ought to do *A*. But I don't think they show that someone who intends to do *A* need not believe that she ought to do *A*. *Before* Zach makes any decision, of course, the carton on the left and the carton on the right are perfectly on a par. But *after* he has made his decision, the carton on the left comes out on top. The fact that he has picked it is now a relevant difference between the two. So if he ought to take one of them, the one on the left is the one that he ought to take.

It is sufficient for this answer that intentions provide reasons—when you intend to do something, that gives you some additional reason to do it, that you did not have before. Call the element of a decision that gives you an additional reason to do it *plumping*. Plumping need only provide you with a very trivial reason, in order to deal with cases of mere picking. So plumping for hiring an assassin to murder your wife is not sufficient to make that what you ought to do. But plumping for the carton on the left *is* sufficient to make that what you ought to do. We do it all of the time. On the view that intention involves the belief that you ought, therefore, Zach is able to intend to take the carton on the left, because he is able to plump for it, and because he understands that his plumping is all that it takes to make a difference between the two. Consequently, though they provide a *prima facie* counterexample, cases of picking present no serious obstacle to the view that intending involves believing that you ought.

8.4.2 The Problem of Akrasia

A harder problem is that of akratic intentions, which seem not only to be possible, but, as Wallace [2001] and others have emphasized, to be subject to the norm of means–end coherence. In cases of *akrasia*, an agent acts contrary to what she believes she ought to do. So they also appear to be counterexamples to the thesis that intending requires believing that you ought. Exactly how to understand what goes on in cases of *akrasia* is quite a large philosophical problem in its own right; I'll confine myself here to four limited observations.

First, it's worth observing that if it is possible to have contrary beliefs, then one possible view is that someone who intends akratically believes that she ought not to be doing what she intends, but also that she ought to. Her situation might be thought to be analogous to that of the man who deep down believes that his wife is cheating on him, but who, due to wishful thinking, manages to convince himself that she is faithful. If the akratic's situation is like this, then though she suffers from a vice that is akin to wishful thinking, she is not, after all, a counterexample to the thesis that intending involves believing that you ought. Still, this is not likely to satisfy most, and so I have three observations about how something like the nature of intention can be defended by appeal to weaker assumptions to which the akratic is not a counterexample.

The first way that we could weaken the view, is by weakening what the intender has to believe. Instead of requiring that he believe that he ought to do it, we could

instead require that he believe, say, that he has *adequate reason* to do it. Then it would follow, presumably, from an appropriate analogue of the subjective ought test, that intenders have *adequate subjective reason* to do what they intend, and hence by a generalization of the transmission principle, adequate subjective reason to do what they believe to be necessary to what they intend. Or we might weaken the content in some other way. Whichever way, it would yield the same style of explanation, with weaker assumptions.

Another way to weaken the required assumption might be to generalize, and allow that an intender need not *believe* that she ought, so long as she *takes it* that she ought. Many authors have claimed, after all, that intentional action requires that you 'take yourself' to have a reason. I'm here assuming that 'taking' amounts to a different attitude than belief—perhaps it is quasi-perceptual. Suppose, then, that intention requires *taking* it that you ought. If that is so, then all that we would need would be a small revision in our test for subjective oughts, in order to have the same explanation as before. We have been supposing that your subjective reasons depend on your beliefs. But suppose that we generalized, and held that your subjective reasons depend on what you either believe or *take* to be true. (To motivate this, we might imagine a Ewing-like case of someone who has no *belief* about what he ought to do, but does *take* it that he ought to do *A*.) Given this broader principle, our explanation would go through as before, simply with a weaker assumption about what intention requires. This proposal, I think, may have some promise at addressing the worry about akratic intentions.

Finally, we could drop the assumption that Zach must have a belief about what he ought to do, altogether, and instead require merely that he have some beliefs which are the sort, if they are true, to guarantee that Zach ought to do it. This weaker assumption would guarantee that when Zach intends to do *A*, he subjectively ought to do it, and hence subjectively ought to do what he believes necessary for it, by the subjective reflection of the transmission principle.

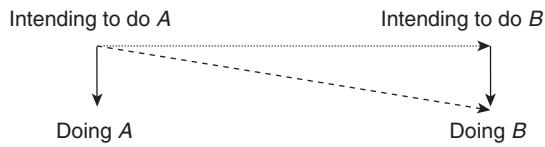
All three of these are ways of weakening the crucial assumption that this view about means–end coherence of intention requires. My point here is not to advocate any of them in particular—that would require having a great deal more to say about *akrasia* than I can in the scope of this paper. My point is rather to show that the essential attractions of the explanation in section 8.3.3 can be preserved under weaker assumptions than premise 2.1. The most I can hope to convince you of here, is that there are grounds for optimism that some combination of these kinds of adjustments can accommodate the problem of *akrasia* in a way that preserves the attractions of this kind of explanation of the norm of means–end coherence, and that the attractiveness of the explanation which results makes it worth at least trying to do so.

8.4.3 Objection: We've Explained the Wrong Thing

The astute reader will have noticed that I set out to explain why there is something wrong with Zach if he does not *intend B*, but what I have actually done, is to explain

why there is something wrong with Zach if he does not *do B*. His belief that *B* is necessary for *A*, on the view I'm outlining, makes it the case that he subjectively ought to do *B*, due to the transmission reflection and the fact that the intention to do *A* entails the belief that you ought (on the unmodified version of the view). So that explains why he subjectively ought to do *B*, but not why he subjectively ought to *intend* to do *B*.

This is true. I haven't yet explained why Zach subjectively ought to intend to do *B*. But before I say how I would explain this, on the present view, allow me to first draw a picture, in order to contrast my style of explanation with one others have claimed.



Everyone should agree that just as there is an important normative relationship between intending to do *A* and intending to do *B*, there is an important relationship between intending to do *B* and doing it, and an important normative relationship between intending to do *A* and doing *B*. On the usual view, this last relationship, which I have drawn with the diagonal arrow, is compound: it is the result of *composing* the other two relationships, each of which is interesting and distinctive.[26] According to this view it is *one* thing to understand the horizontal arrow, and *another* to understand the vertical arrows, and only once we are in a position to understand both, will we be able to put them together in order to understand the diagonal arrow.

Because this view is so common, it can seem surprising that what I have just done, is to have explained the diagonal arrow without first explaining the horizontal one. But having done so, it is natural to reverse the usual order of explanation, and to use our account of the diagonal arrow in order to offer an account of the horizontal one. In fact, doing so is easy. Our explanation so far is this: intending to do *A* makes it the case that you subjectively ought to do *A*. And given this, believing that *B* is necessary for *A* makes it the case that you subjectively ought to do *B*. So by the same principles, believing that intending *B* is necessary for doing *B* makes it the case that you subjectively ought to intend *B*. Just as doing *B* is a means for doing *A*, intending *B* is a means for doing *B*. So from transmission reflection, it follows that if you believe that to do *B* you must intend *B*, then you subjectively ought to intend *B*.

Of course, this does not explain why *anyone* who intends *A* and believes that to do *A* she must do *B* subjectively ought to intend *B*. It only explains this for the special case of people who further believe that to do *B* they must intend *B*. But this is not a bad thing! It is exactly what we need in order to distinguish the cases of *means*, to which the norm of means–end coherence applies, from those of *foreseen side-effects*, which the norm of

[26] For example, see Broome [2001].

means–end coherence does not require you to intend.[27] If you are the Strategic Bomber and intend to bomb the munitions factory which is next to the school, you believe that the only way that you can bomb the munitions factory is if you bomb the school. Since you subjectively ought to bomb the munitions factory, you subjectively ought to bomb the school. But it is not the case that you subjectively ought to intend to bomb the school, because you don't believe that you need to intend this, in order to bomb it. You think that it will simply happen as a result of bombing the munitions factory.

So, to recap: Zach intends to do *A*, and this entails that he has some belief which, if true, guarantees that he objectively ought to do *A*. So it follows from the right-to-left direction of the subjective ought test that he subjectively ought to do *A*. He also believes that to do *A* he must do *B*, and so from the transmission principle for subjective oughts it follows that he subjectively ought to do *B*. And finally, provided that he believes that to do *B* he must intend *B*, it follows from the transmission principle for subjective oughts that he subjectively ought to intend *B*. And this further assumption about his beliefs is appropriate to make, because once we look closely at cases of unintended side-effects, we see that the norm of means–end coherence doesn't really apply in cases in which he doesn't have this belief, anyway. I'll offer some revisions in part 5, but this is the basic idea behind my proposal for how to account for the norm of means–end coherence on intention.

8.4.4 A Temporal Advantage of this Account

One advantage of this account is that unlike many accounts of cases like Zach's, it captures the right relationship to his intention over time.[28] Whatever goes wrong with Zach when he intends to do *A* but has no intention to do *B*, it is not going wrong with Zach immediately after he decides to do *A* and before he has time to think about what is necessary for doing *A*. It only goes wrong if *at no time* prior to the time at which Zach intends to do *A*, does he form the intention to do *B*. This is something that many accounts of cases like Zach's get wrong. For example, according to Wide-scope accounts, Zach is at all times under a special requirement to not be such that he both intends to do *A* and does not intend to do *B*. So it follows from such accounts, that whatever is going wrong with Zach is going wrong with him even before he has had a chance to reflect on what is necessary for *A*. So Wide-scope views seem to predict the wrong results about what happens in Zach's case over time.

The view developed here, however, gets what I think are the *right* results about this case. Since it is only necessary to intend to do *B* at some time prior to the time at which *B* actually needs to be done, there need be nothing strictly wrong with Zach if he has not yet formed the intention to do *B*. Though he subjectively ought to form this intention, he is not *strictly* required to do it right away. Since that seems like the right

[27] I take this moral to be familiar from the work of Michael Bratman, who has pressed it particularly acutely.

[28] Compare Raz [2005b, 5].

result, and other views seem to get the wrong result, or to build in the right result only through stipulation, I think that is collateral evidence that this view about the norms of means–end coherence on intention is worth taking seriously.

8.5 Generalizing

8.5.1 Reconsidering the Subjective Ought Test

I think I have illustrated enough for us to see the appeal of this view. It gives us an answer to the question of *in what sense* there is something going wrong in Zach's case when he does not intend *B*—a problem that the failures of Wide-scope accounts and of the naïve view show is hard. Moreover, the answer that it gives us to the *in what sense* question also leads to an answer to the *why* question that appeals, with one exception, only to plausible independent principles: that there is a distinction between objective and subjective 'oughts' which are related in something roughly like the way indicated by the subjective ought test, and that objective oughts transmit to necessary means. The only other necessary assumption is that an agent who intends to do *A* necessarily satisfies some condition which would give her a subjective reason to do *A*, and in part 4 I suggested that some version of this claim might very well be defensible, though I have not had the space, here, to go into defending some version of this hypothesis in detail. Certainly it is the sort of thing that philosophers have believed for independent reasons.[29]

I've formulated the account so far by appeal to the subjective ought test as formulated in section 8.2.1. This has been because that principle has been accepted by many philosophers, including philosophers with different views about, for example, the priority between the subjective and objective 'ought' relations, and because it is easy to motivate by consideration of initial examples like that of Bernie. But in fact, however, I believe that the subjective ought test is correct only in spirit. I'll now explain why I do not believe that it can actually be correct. Then in section 8.5.2 I'll briefly defend what I think is right, instead, and use that to show that the essentials of the account of means–end coherence are unaffected. Finally, in section 8.5.3 I'll argue that the modified account has an advantage not shared by most existing accounts of the norm of means–end coherence on intention.

According to the subjective ought test formulated in section 8.2.1, if Yves believes that he objectively ought to do *A*, then Yves subjectively ought to do *A*. This is because, in believing that he objectively ought to do *A*, he believes something which guarantees (trivially) that he objectively ought to do *A*. From the same test, it also follows that Bernie subjectively ought to take a sip, conditional on our assumptions that Bernie believes that his glass contains gin and tonic, and that if it contains gin and tonic, then he objectively ought to take a sip. But now suppose that Yves *is* Bernie and that *A is* the

[29] Compare, for example, Davidson [1978] and Tenenbaum [2007].

action of not taking a sip. That is, imagine that Bernie believes that his glass contains gin and tonic, but also, for some reason, believes that he ought not to take a sip. Our test as so far formulated predicts that Bernie subjectively ought to take a sip, and also that he subjectively ought to not take a sip.

This seems bad. It should not turn out that Bernie subjectively ought to do both things. Intuitively, the problem is that what Bernie subjectively *ought* to do should turn on *all* of his beliefs, not simply on one or two. What he subjectively *ought* to do should turn on what happens when you put the effects of all of his beliefs together.

But on the other hand, tests for the relationship between objective and subjective 'oughts' that turn on the whole sum of an agent's beliefs are subject to other kinds of counterexamples. For example, Jacob Ross considers the following sort of case:[30] Wynn has the opportunity to choose one of three envelopes set in front of her. Whatever it contains, she will be able to keep, and she will not get what is in the other two envelopes. She believes that the first envelope contains two hundred dollars, that one of the other envelopes contains three hundred dollars, and that the other contains nothing. And she considers it equally likely that the three hundred dollars are in the second envelope as in the third. According to Ross, she subjectively ought to take the first envelope, because it has the highest expected value, given her beliefs. But if her beliefs are all true, then one of the following has to be the case: either there is more money in the second envelope than in the first, or there is more money in the third envelope than in the first. Either way, it would be a mistake to take the first envelope. The thing she *objectively* ought to do, is the one that will actually be best. It follows that it can't be true that Wynn subjectively ought to do whatever it would be the case that she objectively ought to do, were all of her beliefs true. *Incompleteness* in her beliefs can play a role in fixing what she subjectively ought to do.

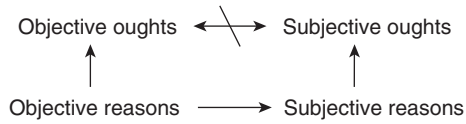
As I anticipated in part 4, the three-envelope problem is also a serious obstacle to the thesis that intending to do *A* requires believing that one objectively ought to do *A*. When Wynn chooses an envelope, she does so despite knowing full well that she objectively ought to choose one of the others—she simply doesn't know which.

Finally, I'm independently wary about the tests which posit a direct relationship between the objective and subjective 'oughts'. This is because I believe not only that there are objective and subjective 'oughts', but that there are objective and subjective reasons. And it seems to me that objective oughts are related to objective reasons in the same way that subjective oughts are related to subjective reasons. Finally, I hold that oughts are to be accounted for in terms of reasons, rather than conversely.[31] You ought to do what the balance of your reasons favors, I hold. It follows from these views that there can't be a direct reductive relationship between the objective and subjective 'oughts' that would validate anything like our tests. If objective oughts are to be

[30] Ross [2006] and Parfit [2011] also discuss such a case. The original version of such a case is attributed to a footnote in Regan [1980].

[31] The converse view is articulated by, for example, Toulmin [1950] and Broome [2004].

accounted for in terms of objective reasons, and subjective oughts are to be accounted for in terms of subjective reasons, then neither objective oughts nor subjective oughts can directly be accounted for in terms of the other. The following picture shows how I think things work (the arrows, read from head to tail, represent 'is analyzed in terms of'):



I can't, of course, defend each of these views here, but I think they are relevant to a complete appreciation of the grounds for thinking that objective oughts and subjective oughts might be directly related.

8.5.2 The Correct View, Framed in Terms of Reasons

All of these problems are solved, I think, by the idea that it is objective and subjective *reasons* which are directly related, and related by something very much like the subjective reason test formulated in section 8.2.3. On this view, though it does not follow immediately from the fact that Bernie believes that his glass contains gin and tonic that he ought to take a sip (for he might also believe that he ought not), it does follow that he has a subjective *reason* to take a sip. That is why, if other things are equal, it *will* be the case that he subjectively ought to take a sip. For what he subjectively ought to do, is simply whatever he has most subjective reason to do. So this solves our first problem.

It also addresses Ross's kinds of case. In Ross's case, if Wynn's beliefs are all true, then it is not the case that she objectively ought to take the first envelope. It is either the case that she objectively ought to take the second, or that she objectively ought to take the third. On my view, this is because what she objectively ought to do is a matter of *all* of her objective reasons, which include the following: the fact that there is $200 in the first envelope is a reason to take it, and the fact that there is $300 in the second envelope is a reason to take it. But since the second reason is a weightier reason than the first, and there are no other relevant reasons, she objectively ought to take the second envelope. The same thing goes, if the $300 is in the third envelope instead of the second.

But what Wynn subjectively ought to do, on my account, is a matter of the weight of her subjective reasons. And those include the following: her belief that the first envelope contains $200 gives her a reason for her to take the first envelope, her belief that the second envelope is 50% likely to contain $300 gives her a reason for her to take the second envelope, and her belief that the third envelope is 50% likely to contain $300 is a reason for her to take the third envelope. But since she has no view about which of the second or third envelope contains the $300, she has no further belief to figure as a reason. Now which of these subjective reasons is weightiest? Well, suppose they are all true. The fact that the first envelope contains $200 would seem to me to be a weightier

reason for her to take it, than the fact that the second envelope is 50% likely to contain $300. So just based on *those* two reasons, the thing for her to do would be to take the first envelope. Assuming that subjective reasons are as weighty as their corresponding objective reasons would be, therefore, it follows that Wynn's subjective reason to take the first envelope is the weightiest.[32]

Of course, if these beliefs are true, then some further fact must also be true, to the effect that the second envelope either does, or does not, contain the $300. But since Wynn does not have a belief about that, it does not figure among her subjective reasons by way of the converse reason test. So I conclude that what Wynn subjectively ought to do, as Ross insists, is to take the first envelope. We actually predict this result, if we take the direct connection to hold between objective and subjective reasons, rather than between objective and subjective oughts.

Moreover, switching from oughts to reasons dissolves the problem raised by the three-envelope problem for the thesis that intending involves believing you ought. True, due to the three-envelope problem, it is not tenable to claim that intending to do *A* requires believing that you objectively ought to do *A*. But it may still require believing, or taking it that, you have a sufficiently weighty objective reason to do *A*.

Finally, these revisions still suffice for the account of Zach's case, basically substituting 'reason' everywhere in our earlier explanation for 'ought'. On the assumption that intending to do *A* entails believing that you have a reason to, it follows from the right-to-left direction of the subjective reason test that if you intend to do *A*, then there is a subjective reason for you to do *A*. What we then need is a subjective reason transmission principle, which I suggest that we can derive from the subjective reason test together with the generalization of the objective ought transmission principle to the case of reasons. Here I'll state what I take to be a special case:

reason transmission: If *X* has an objective reason to do *A* and to do *A X* must do *B*, then *X* has an objective reason to do *B* of equal weight to *X*'s objective reason to do *A*.

In the same way as for the transmission principle for objective oughts, we can use this to derive its subjective reflection. (Actually, the derivation requires principles which include reference to weights, which is stronger than the principles stated in section 8.2.2.) I'll just state the transmission principle that I think is plausible:

[32] Officially, my views about the weights of reasons are more complicated than this. See Schroeder [2007a, chapter 7]. But I don't think that I am committed to anything which would make this principle fail in the simple cases that we are considering. A helpful referee also encourages me to note that determining the weights of subjective reasons, as I do in an informal and intuitive way throughout this section and the next, is going to be a complicated matter—particularly because they can derive from sets of beliefs which are not themselves consistent. There is unfortunately insufficient space to take this issue up in sufficient depth, here, so I'll reserve it for a future occasion. My goal in this paper has been to stress the attractions of this basic idea, and there remains considerable work to be done both in defending the required thesis about the nature of intention, and in articulating and precisifying the intuitive claims I'm making about the weight of reasons, both objective and subjective.

reason trans reflect If *X* has a subjective reason to do *A* and *X* believes that to do *A* she must do *B*, then *X* has a subjective reason to do *B* of weight at least as great as *X*'s subjective reason to do *A*.

It follows from this principle that Zach has a subjective reason to do *B* that is at least as weighty as whatever subjective reason he has on account of intending to do *A*. This accounts, I claim, for the *stringency* of the norm of means–end coherence, the feature that I claimed in section 8.2.2 was central to what is difficult about this case. Means–end coherence is *stringent* because when you believe that the means is *necessary* for the end, the full force of your subjective reason to pursue the end transmits to your subjective reason to pursue the means.

8.5.3 A Final Advantage: Generalizing from Necessary Means

Finally, I want to show how this allows us to bring out a final important feature of this account. Just as there is some interesting relationship between intending ends and intending means which are necessary to those ends, so also there is a more general connection between intending ends and intending non-necessary means that nevertheless *facilitate* those ends—which help to bring them about. But most existing views about the norm of means–end coherence on intention trade essentially on features of the paradigm case, in which the agent's means–end belief is that the means are *necessary* for the end.

For example, on Harman's [1976] account, on which many other accounts have since been modeled, you ought to intend the means you believe are necessary, because the intention to do *A* entails the belief that you will do *A*. And if you believe that you will do *A* and that if you do *A* you will do *B*, then theoretical rationality requires the belief that you will do *B*, and it is because you are required to believe this, that you are required to intend the means (the further details aren't essential for the point that I want to make). This models the means–end relation on *modus ponens*. But that seems to require a *deductive* relationship between end and means. It makes it hard to see how it will generalize to the more general case in which Zach has beliefs about what would facilitate his end which he doesn't take to be strictly necessary for it. Similar features are shared by the accounts offered by Wallace [2001], Broome [2001], and Setiya [2007a].

On the account offered here, on the other hand, it is easy to see why things should generalize, and why we will get the intuitively right results when we generalize. This is because the transmission principle for reasons that I have stated is really only a special case. In general:

general reason transmit: If *X* has an objective reason to do *A* and *X*'s doing *B* would facilitate her doing *A*, then *X* has an objective reason to do *B* of weight at least proportional to *X*'s objective reason to do *A*, and to how well her doing *B* would facilitate her doing *A*.[33]

[33] Compare Raz [2005a]. I take these this principle to be intuitive, even though I have no way of making the 'proportionality' claim precise.

And this, of course, yields, together with suitable versions of our tests for subjective reasons and their weights:

subjective version: If *X* has a subjective reason to do *A* and believes that her doing *B* would facilitate her doing *A*, then *X* has a subjective reason to do *B* of weight at least proportional to *X*'s subjective reason to do *A*, and to how well she believes her doing *B* would facilitate her doing *A*.

But by now it should be obvious that we can use such a principle in order to explain subjective reasons for Zach to take what he believes to be the facilitative means to his intended ends. Moreover it predicts, correctly, that these subjective reasons will not be as weighty as the subjective reasons to take the means he believes to be necessary. And so it allows that this case does not exhibit the stringency of the necessary means case. All of these predictions are intuitively correct.

In my view, it should be a constraint on an adequate account of Zach's case—of the so-called norm of means–end coherence—that it both explain the special stringency of what is going on when Zach's belief is that some means is necessary for his end, and also generalize to explain what is going on in cases where he has beliefs in non-necessary but facilitative means. So I take it to be an important and distinctive virtue of this account that it looks like it can do this correctly, and by appeal to relatively weak assumptions.

To reiterate: what we need for this account is some general relationship between objective and subjective reasons along the lines of the subjective reason test formulated in section 8.2.2, along with corresponding claims about those reasons' weights. Such a principle is motivatable entirely independently of the problem about means–end coherence, and I argued in sections 8.5.1 and 8.5.2 that its correct version withstands problems for less careful formulations. We also need to appeal to something in the neighborhood of general reason transmit, which looks like a highly plausible and totally general principle, at least something like which we should independently accept. And finally, we need to accept some hypothesis about the nature of intention—for example, that intending to do something requires taking yourself to have a reason to do it.

I haven't argued for this final hypothesis. But I'm definitely convinced that the considerations adduced here show that it would be an explanatorily powerful assumption, if it were true. I know of no other account either of *in what sense* there is something wrong going on in Zach's case, nor of *why* this is so, which is entirely happy. And this account not only avoids the pitfalls for others, but deals correctly with the way that the norm of means–end coherence applies over time, and generalizes in the right kind of way to accommodate merely facilitating means. So the hypothesis seems well worth exploring further than I have been able to do, here.[34]

[34] The central ideas of this paper were formulated during Michael Bratman's presentation to the 2006 conference on Practical Reason at Bowling Green State University. My apologies to him for being distracted during the talk. I also owe thanks to Bratman, Doug Lavin, and Niko Kolodny for listening to my incoherent formulations that weekend, to Joseph Raz, to Kieran Setiya's enlightening paper, to conversations with Jacob Ross, to very helpful comments from Jamie Dreier, Steve Finlay, and Jacob Ross, to an excellent and insightful referee, and to an audience at Georgetown University.

Part 4

The will is therefore not merely subject to the law, but subject in such a way that it must be considered as also *giving the law to itself* and only for this reason as first of all subject to the law (of which it can regard itself as the author).

[Kant 2002, 232 (4:431)]

9

The Hypothetical Imperative?

9.1 Introduction

The contrast between hypothetical and categorical imperatives is central to Kant's ethical theory.[1] According to Kant, all imperatives present a practical kind of objective necessity, but with hypothetical imperatives that necessity is *based on a presupposition*, while with categorical imperatives it is not. Hypothetical imperatives commend taking the necessary means to a certain class of agents—those who have the end. *If* you have the end, *then* you ought to take the necessary means. But many contemporary commentators on Kant disagree. According to them, Kant believes in a Hypothetical Imperative that is not based on any presupposition; instead, it has a disjunctive *object*. It tells everyone to "[t]ake the necessary means or else give up the end" [Hill 1973, 436]. They say that, "[t]aken strictly, it [the hypothetical imperative] counsels us *either* to take the means *or* to give up the end" [Darwall 1983, 16].

One of the ways in which the problem can be seen to arise[2] is to follow deontic logicians in interpreting 'ought' as a sentential operator, on the model of 'necessarily' and 'possibly'. In deontic logic, 'ought' assigns *practical*, as opposed to *alethic* or *metaphysical* necessity to propositions. So, in the generic statement of hypothetical imperatives—'if you will the end, then you ought to take the necessary means'—there are two salient choices for how to interpret the scope of the 'ought', since there are two salient sentential clauses in which it figures. One choice is to give it scope merely over the consequent of the conditional. Let us call this the *Consequent Scope* reading of the

[1] Page references in the text are to the Prussian Academy edition of Kant's works, by volume and page number, except where indicated otherwise.

[2] 'Ought' is probably not a sentential operator (see Chapter 7, this volume), and Kant certainly didn't think that it was. Kant held that 'ought' expresses a relation between someone who ought to do something, and an *action*—not a proposition—which she ought to *do*. Imperatives, that is, "say that *something would be good to do* or to refrain from doing, but they say it *to a will* that does not always" do it [4:413, my italics]. Fortunately, however, we don't need this assumption in order to get this dispute going. Kant never actually says, 'if you will the end, then you ought to will the necessary means.' He only says things like, "[a] hypothetical imperative thus says that an action is good for some purpose, either possible or actual" [4:414–415]. So we don't need the thesis that the sentence is actually ambiguous—we only need ideas like those of Hill and Darwall for what hypothetical imperatives might require.

conditional. The other choice is to give it scope over the entire conditional. Let us call this the *Wide Scope* reading.[3]

Consequent Scope: $\forall x \forall e$ (if x wills end e, then x ought (x takes the means to e))
Wide Scope: $\forall x \forall e$ (x ought (if x wills end e, then x takes the means to e))

It is important to notice *two* obvious differences between Consequent Scope and Wide Scope. Because the difference between them derives from whether 'if x wills end e' qualifies which x ought to do something, or whether it qualifies what the x ought to do, Wide and Consequent Scope disagree not merely about what kind of thing hypothetical imperatives enjoin us to do. They also differ with respect to *who* they enjoin to do it. According to Consequent Scope, hypothetical imperatives enjoin *those with the end* to take the necessary means. Their mandate is, as Kant puts it, "based on a presupposition," and they don't mandate it to those who don't satisfy the presupposition. But according to Wide Scope, the Hypothetical Imperative is not based on any presupposition. It applies to everyone unconditionally, no matter what they are like. It sounds, in short, very much like a special kind of *categorical* imperative.

The dominant view of Kant on hypothetical imperatives is that Kant accepts hypothetical imperatives only with a Wide Scope reading.[4] But in this paper I'll explain why this is a clear mistake, at least with respect to the Kant of the *Groundwork* and the second *Critique*. I'll point out, in fact, that the only serious motivation for this interpretation of Kant stems from considerations which are entirely extra-textual,[5] and I'll argue that Kant's view plausibly possesses other resources to deal with these considerations.

[3] Stephen Engstrom [1993, 408] classifies the central options slightly differently. He seems to think that there is only one possible version of the Consequent Scope view, and it is that it is to be restricted to ends that are 'set by reason' [408], and hence obligatory. He calls this view the 'material' reading of hypothetical imperatives, and contrasts it with the 'formal' reading, which is essentially Wide Scope. His 'material' reading of hypothetical imperatives, however, is itself merely a special case of the Consequent Scope reading—it is *possible* to hold a Consequent Scope view which applies even to ends that are not 'set by reason'. This non-exhaustive classification of views might appear to be a slight problem for Engstrom, since his argument in that paper is supposed to pose a dilemma for one of Allison's views. But as we'll see in section 9.8, I do hold that a version of the 'set by reason' view is in fact that way that Kant's Consequent Scope view works.

[4] Hill [1973] is the classic and one of the clearest statements of this interpretation and its motivation. Hill also later reiterates his position [Hill 1989, 365]. I take it that Rawls would agree, since Hill cites his lectures as inspiration [Hill 1973, 429]. Korsgaard endorses Hill's reading [Korsgaard 1997b, footnote to 234], and offers two detailed accounts of what Kant's arguments were for this view at different points in his career, which I'll consider in part 2 [Korsgaard 1997b, 238–240, 245–247]. Barbara Hermann follows Hill at least as far as talking about 'the Hypothetical Imperative' in the singular [Hermann 1993, 144, 214–215]. Stephen Darwall [Darwall 1983, 16, 47] and Jean Hampton [Hampton 1998, 144, 165] certainly commit to the view, and I suspect that many if not most contemporary Kantian thinkers in ethics are guilty of making the same mistake. H. J. Paton is extremely unclear about how he uses the word 'conditioned' in his *The Categorical Imperative* [Paton 1947], but he clearly thinks, like Hill, that individual hypothetical imperatives are 'conditioned' on the Hypothetical Imperative, which is in turn not 'conditioned' on ends at all, but only on the moral imperative. See section 9.9.

[5] Greenspan [1975], Gensler [1985], Hampton [1998], Broome [1999], Chapter 7, this volume. Simon Blackburn [Blackburn 1998] also endorses this reading of hypothetical imperatives. Decision theory, conceived as a theory of instrumental reason, is probably also of the Wide Scope variety, and for similar reasons [Dreier 1996].

It will be a consequence of my interpretation that, *contra* interpreters like Korsgaard and Hill, Kant never seriously took there to be a Hypothetical Imperative, capital-H, capital-I, an independent and fundamental objective principle of practical reason of which hypothetical imperatives are the expression, in the way that categorical imperatives are expressions of the Categorical Imperative.

The structure of the paper is simple. In the remainder of part 1, I present two arguments in favor of the Consequent Scope reading of hypothetical imperatives, either of which I think ought to be conclusive. Then in part 2 of the paper, I turn to Kant's famous argument that hypothetical imperatives are analytic. I argue that given Kant's conception of analyticity, there is only one viable way of interpreting this argument, and that on this interpretation the argument only supports Consequent Scope. Finally, in part 3 of the paper I turn to consider the motivation for the Wide Scope interpretation. I note that while my arguments are textually based, this motivation derives from an extra-textual philosophical problem faced by contemporary interpreters of Kant. And I note that if my interpretation of the analyticity argument is correct, then Kant already has a solution to this problem.

9.2 Wide Scope and Categoricity

Among imperatives, as I noted above, Kant distinguishes between those which are *categorical*, and those which are merely *hypothetical*. Categorical imperatives are addressed to all finite rational agents, and are not also hypothetical imperatives. Hypothetical imperatives, according to Kant, address themselves only to those who have certain ends. This doesn't mean that they don't address themselves to all finite rational agents. The imperative of prudence, according to Kant, *is* addressed to all finite rational agents. *Everyone* ought to act to ensure their own happiness. But this is only because an agent's happiness is determined by what her ends are—it is whatever would fulfill them. So no matter what ends an agent wills, acting to ensure her happiness is taking the means to her ends. And so she ought, because of the ends that she wills, to act to ensure her happiness [4:415]. Real categorical imperatives, in Kant's sense, aren't like this. There need be and can be *no* explanation in terms of agents' ends as to why a real categorical imperative is addressed to everyone [4:416]. We cannot, as Kant says in my epigram, base it on any presupposition about the agents to whom it is addressed.

According to Kant, moreover, hypothetical imperatives are much easier to understand than categorical ones. We can show by means of a simple analysis of the concept of willing an end that hypothetical imperatives are possible. But we need a substantive, synthetic argument, and an appeal to the autonomy of rational agents, in order to discover how categorical imperatives are possible. Moral imperatives, according to Kant, statements about what people ought, morally, to do, are addressed to all finite rational agents. And even though such imperatives aren't *thereby* categorical—the imperative of prudence, after all, is addressed to everyone, but isn't categorical—Kant doesn't see how the *content* of moral imperatives could be derived in the same way as

his imperative of prudence, since they are not, like that imperative, seemingly 'empty'. So he concludes that they must in fact be categorical. And so moral imperatives turn out to be much harder to understand than hypothetical ones, in part *because* they are categorical. This is what he says at [4:419] of the *Groundwork*, the passage I took as my epigram.

All of this makes Wide Scope initially very hard to credit as an interpretation of Kant's account of hypothetical imperatives. For according to that interpretation, if people with playing the cello well as their end ought to practice, it is only because everyone ought to ensure that if he has an end, then he takes the necessary means. That is, according to Wide Scope, hypothetical imperatives depend on the truth of an *ought* which applies to everyone. Moreover, this *ought* cannot in turn be explained by appeal to everyone's ends, in the way that the imperative of prudence was. For it is itself supposed to *explain* how *any* 'ought'-statements can follow from facts about ends. So it must be an imperative which is *categorical* in Kant's sense. And thus, contrary to Kant's claim that what makes moral but not hypothetical imperatives puzzling is that they alone presuppose a *categorical* imperative, it turns out that even hypothetical imperatives presuppose a categorical imperative.

Jean Hampton, who seems to take the Wide Scope approach to giving an account of practical reason to be the only possible view, bites the bullet quite explicitly:

> Kant's position on the nature of hypothetical imperatives must be construed (**contra his explicit wishes**) such that understanding the bindingness of a hypothetical imperative is no easier than understanding the bindingness of a categorical imperative. My interpretation cannot save Kant's belief that the former is more straightforward than the latter; **indeed, my argument is that Kant's belief is wrong**. The only way to analyze Kant's analyticity claim is to do so in a way that locates in hypothetical imperatives the same mysterious objectivity that attends the categorical imperative. Even more strikingly, I have argued that the force of hypothetical imperatives is dependent on, and *is at least in part constituted by*, the force of some antecedent categorical imperative that is in part definitive of instrumental rationality.[6]

As I've noted, if Kant intends hypothetical imperatives to be understood on the model of Wide Scope, Hampton is right that they presuppose a categorical imperative. But given that it is at the absolute center of Kant's practical philosophy that hypothetical imperatives are merely analytic, but we cannot show that a categorical imperative is even possible, except by synthetic means, this really ought to suggest to us that we should think rather harder about whether to interpret him on that model.

[6] Hampton [1998, 165–166]. The boldface type is my addition, but the italics are her own emphasis. Hampton certainly seems to be offering a Wide Scope reading: "[s]o although in fact there are agents who desire the end, but not the means to the end, this principle says that they ought not to do so, and will be condemned as irrational to the extent that they do so" [Hampton 1998, 133]. But in fact, much of her discussion is ambiguous or unclear on this point, and what really seems to be important for her argument are the claims I've put in boldface, and the last quoted sentence. So long as hypothetical imperatives must derive from some categorical imperative, Hampton ought to be happy, even if that categorical imperative isn't of the Wide Scope variety. I discuss another possible interpretation of what Hampton might actually have been thinking in Chapter 1.

If it is possible to use analytic means to show the possibility of hypothetical imperatives, and the existence of hypothetical imperatives entails the existence of a categorical imperative, then it turns out that it's possible to demonstrate the existence of a categorical imperative by analytic means after all. So it becomes extremely puzzling why Kant would have thought that it was the very *categoricity* of moral imperatives which meant that we needed a synthetic argument and an appeal to the autonomy of rational agents in order to see how they are possible. It must have been entirely different reasons which made Kant think that the moral imperative was more puzzling or difficult to understand. Thomas Hill's 'The Hypothetical Imperative' [Hill 1973], one of the classic sources for the Wide Scope interpretation of Kant, is mostly devoted to defensive maneuvering on just this front.

It's not that the considerations which Hill offers are irrelevant. On the contrary, he highlights a number of features of the moral imperative *besides* its categoricity which distinguish it from the Wide Scope version of the hypothetical imperative. But none of this gets past the point that Kant obviously thought that *categoricity* was *also* an important distinguishing characteristic of moral imperatives, and the Wide Scope interpretation makes this out to be complete nonsense, and obviously so.

9.3 The Problematic Assertoric Distinction

In the *Groundwork* and in a set of lectures given at about the same time the record of which we owe to Mrongovious, Kant draws a distinction that makes no sense if his view is Wide Scope, but which is enormously important, if his view is Consequent Scope. The distinction is that between two kinds of hypothetical imperatives—ones which are *assertoric*, and ones which are merely *problematic*. Assertoric hypothetical imperatives are ones whose end is actually given, while problematic hypothetical imperatives are ones in which someone doesn't actually have the end. Kant clearly thinks that this is a philosophically important distinction,[7] but it is hard to make sense of this on a Wide Scope reading. On the Wide Scope reading, after all, *whether or not* you have the end, you ought to ensure that *if* you have it, you take the necessary means. The fact that you actually *have* the end doesn't actually make any difference as to what you ought to do. But on the Consequent Scope reading, your having the end makes *all the difference*. If you have the end, then there is something that you ought to do, that it wouldn't have been the case that you ought to do, if you didn't have the end. We can *assert* of you, that you ought to take the means, for you now fall under the application class of the *ought*, while before you did not.

[7] Perhaps Kant does sometimes make distinctions more in the interest of taxonomy than philosophical importance. But it's extremely hard to imagine that this is the case in this instance. For one, it's hard to even see what the distinction *is* on the Wide Scope reading, much less why it is important, or why Kant's terminology for it makes sense.

Very interestingly, the 'assertoric'/'problematic' distinction is one that Kant later rejects, in a footnote to his original, unpublished introduction to the *Critique of Judgment*. There he writes that

> [t]his is the place to correct a mistake I made in the *Foundations of the Metaphysics of Morals*: having said there that imperatives of skill command only conditionally—namely, under the condition of merely possible, i.e., *problematic* purposes—I called such practical precepts problematic imperatives. **But in fact this expression is contradictory**. I should have called them technical, i.e., imperatives of art. [20:200, boldface added for emphasis]

Here Kant claims to eschew his earlier distinction on the grounds that strictly speaking, the phrase "problematic imperatives" is actually contradictory. On the Consequent Scope view, but only on that interpretation, it is easy to see why. Since the antecedent of a problematic imperative is not satisfied, strictly speaking there is *no* imperative—only a hypothesis about what imperative there *would* be, if things were different. Since Kant himself claimed to hold that this phrase was incoherent, that highly suggests that he did hold the Consequent Scope view.[8]

9.4 The Analyticity Argument at [4:417]

Kant finds hypothetical imperatives less puzzling than categorical ones, because he thinks that they are *analytic*. And according to Kant, what makes a claim analytic is that its predicate is contained in a correct analysis of its subject. So, for example, it is analytic that bachelors are unmarried, because the proper analysis of 'bachelor' is 'unmarried man'. So saying that bachelors are unmarried is saying that unmarried men are unmarried.[9] Let us, therefore, try to separate Wide Scope and Consequent Scope into subject–predicate form, in order to see how an argument that each was analytic would have to go. Kant himself, after all, does provide such an argument, in the *Groundwork*, at [4:417]. It would therefore be worth our while to take a look at the form that that argument actually takes, and the form that it would have to take, in order to be an argument for Wide Scope or for Consequent Scope.

According to Wide Scope, every agent has the following property: she ought to ensure that if she has an end, she takes the necessary means. According to Consequent Scope, everyone who has an end has the following property: she ought to take the necessary means to it. 'Agent' is the subject of Wide Scope. So Wide Scope would only be

[8] One might question whether an obscure footnote from an introduction that Kant himself chose not to publish should be given very much weight. But it is worth noting that after writing this introduction, Kant insists nearly uniformly on using the term 'technical' to refer to hypothetical imperatives, and abstains from using the term 'hypothetical'—a marked departure, if Kant weren't serious about what he says in this passage.

[9] It is important to note the possibility that Kant has a different conception of analyticity in the practical domain than of analyticity in the theoretical domain, as in the first *Critique*. I'll address this question in section 9.6.

analytic if 'ought to ensure that if she has an end she takes the necessary means to it' could be analyzed as a constituent of 'agent'. So it looks like an argument for the analyticity of Wide Scope would have to start by analyzing 'agent' and finding 'ought to ensure that if she has an end she takes the necessary means to it'. But Kant's analysis isn't an analysis of 'agent' at all. On the contrary, "[t]he imperative derives the concept of actions necessary to this end from the concept of *willing the end*" [4:417, italics added].

Now there are a couple of different ways of reading what is going on at [4:417] of the *Groundwork*. According to folklore, it is that 'wills the necessary means' is part of the concept of 'wills the end'. *He who wills the end wills the means*. But this isn't what Kant actually says. If this were true, as Korsgaard is fond of pointing out [Korsgaard 1997b, 229–230], then it would be impossible to will an end and not will the necessary means, and if so, there would be no sense in saying that one who wills the end *ought* to will the necessary means. This is just a point of which Kant himself is fond—imperatives are only *imperatives* insofar as it is possible to fail to conform to them [4:413]. If it's not possible to will the end and fail to will the means, then neither a Wide Scope nor a Consequent Scope hypothetical imperative will really be an imperative.

As Hill points out, what Kant says at [4:417] is only that "[w]hoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his actions and that lie in his power," or that "[t]he imperative derives the concept of actions necessary to this end from the concept of willing the end." From the latter quotation, it seems clear that Kant *is* analyzing the concept of willing the end, but from the former that what he gets is not "wills the necessary means," but "wills, insofar as he is rational, the necessary means," or simply "ought to will the necessary means."

If this is the right interpretation of the passage, then this fits the Consequent Scope interpretation, but not the Wide Scope interpretation. For Wide Scope to be the correct interpretation, Kant would have to be analyzing instead the concept 'agent'. I'm not arguing, here, that Wide Scope couldn't be analytic, on some conceptions of analyticity. It clearly can be. On some conceptions of analyticity, something is analytic if it is accepted by every speaker of the language competent with the terms involved. Someone might think that Wide Scope is so uncontroversial as to be analytic in this sense. I don't think it is this uncontroversial,[10] but someone might. Nor am I arguing even that Wide Scope couldn't be analytic on Kant's conception of analyticity. It could be thought that there is an analysis of 'agent' from which we get 'ought to ensure that if she wills the end, she takes the necessary means to it'. I don't think there is such an analysis, but that's not what I'm arguing, here. What I'm arguing, is that this is not what is going on in Kant's *actual* supposedly analytic argument for the possibility of hypothetical imperatives. What is going on in this argument is that he is analyzing the concept of willing an end, and finding 'ought to will the necessary means'. This particular argument only supports Consequent Scope, and not Wide Scope.

[10] Chapter 7, this volume.

9.5 Korsgaard on [4:417]

Admittedly, despite Kant's own words: "[t]he imperative derives the concept of actions necessary to this end from the concept of willing the end," others have supposed that the analysis in the passage is really an analysis of something else. Christine Korsgaard, for example, seems to think that it is an analysis of 'rational agent', and her proposal for how to understand the passage is worth considering.[11] According to her interpretation, in "rational agent" we find the concept, "if she wills the end, she takes the necessary means."

> The model suggests that the normativity of the *ought* expresses a demand that we should emulate more perfect rational beings (possibly including our own noumenal selves) whose own conduct is not guided by normative principles at all, but instead describable in a set of **logical** truths. [Korsgaard 1997b, 240, boldface added]

So, according to Korsgaard, 'a rational agent takes the necessary means to her ends' turns out to be analytic, because 'takes the necessary means to her ends' turns out to be part of the analysis of 'rational agent'. Indeed, one of Kant's formulations in this very passage is, "[w]hoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his actions and that lie in his power," and Korsgaard's interpretation gives us a reasonable interpretation of this claim, standing by itself.

But this does not yet give us a statement of Wide Scope. To get Wide Scope as a result of analysis, we would have to analyse 'agent'. But as I interpret Korsgaard, she is suggesting that this is the first step toward deducing Wide Scope. According to Korsgaard, as I understand her, Kant thinks that from this it *follows* that anyone, rational or not, ought to emulate the rational agent, by making sure that she does so, as well.[12] So on my reconstruction of Korsgaard's view, she holds that Kant is not deriving Wide Scope directly, by means of an analysis, but merely claiming that it *follows* from an analytic truth. And that is how, according to Korsgaard, the passage at [4:417] gives us an analytic argument for Wide Scope.

Unlike most philosophers in the twentieth century, however, Kant doesn't seem to have a conception of analyticity on which the consequences of analytic truths are also analytic. On Kant's conception, the analytic truth itself must have its predicate

[11] It is important to note that Korsgaard offers *two* interpretations of Kant's argument for the possibility of hypothetical imperatives. One is the argument that she finds at [4:417], the passage which I am now considering. This interpretation is at Korsgaard [1997b, 238–240]. But since Korsgaard believes that this is not Kant's considered view of the matter, but only a relic of his pre-critical rationalism [1997b, 239], she also has a view about his mature account. This is the final view that she offers in her [1997b] and which she alludes to at [1997a, xvii], and I will consider it as well.

[12] Korsgaard also thinks that this kind of argument is a relic of Kant's precritical rationalism, and doesn't represent his mature view: "[A] perfectly rational being *would* take the means to his ends, therefore I *ought* to take the means to my ends. The model suggests that the normativity of the *ought* expresses a demand that we should emulate more perfect rational beings" [Korsgaard 1997b, 239–240]. Notice that what the (imperfectly rational) agent ought to do is to emulate the perfectly rational one. And the perfectly rational one is this way: she *either* takes the necessary means to an end, *or* she doesn't will it. So an imperfectly rational being ought to be *like that*. And so account of hypothetical imperatives derived is Wide Scope.

contained in its subject. Still, granting Korsgaard two assumptions will get her interpretation going. All that she needs is (1) allowing Kant a more generous account of analyticity, so that analytic truths are those whose predicate is contained in their subject *or else* which follow from such claims, and (2) the claim that it is analytic that any agent ought to do what *rational* agents do. If we grant these assumptions, then Korsgaard's interpretation is one on which the argument, even though it is not an analysis of 'agent' turns out to successfully make Wide Scope analytic.

But I don't see why we should prefer this reading to my own. Kant himself tells us that it is an analysis of willing an end, and *contra* Hill,[13] the argument can be an analysis of willing an end *without* being the analysis that this concept contains the concept, 'wills the necessary means'. It might, as I've suggested, be an argument that the concept of willing an end contains the concept, 'ought to will the necessary means'. And in any case, this alternative reading only works, if we broaden Kant's conception of analyticity, and I know of no independent reason to do so. Moreover, it requires a further argument that it is analytic that any agent ought to do what a rational agent would do, something that even Korsgaard thinks is merely a relic of Kant's precritical rationalism. The most natural way to understand what is going on at [4:417] is to agree that it is what Kant tells us is going on: "[t]he imperative derives the concept of actions necessary to this end from the concept of willing the end." And that means that it can only be an argument for Consequent Scope.

9.6 Korsgaard on Kant's Mature View

So *holding fixed* Kant's account of analyticity, it seems that there is only one viable interpretation of the analyticity argument at [4:417]. And this interpretation forces us to conclude that the argument is an argument for Consequent Scope, rather than for Wide Scope. It might be thought, however, that Kant had a quite different conception of analyticity in the practical domain, than in the theoretical one, as in the first *Critique*. And in fact, this interesting view seems to be presupposed by Christine Korsgaard's preferred version of Kant's account of hypothetical imperatives.[14]

Korsgaard says that "willing the means is conceptually contained in willing the end" [Korsgaard 1997a, xvii], but she also claims that it is possible to will the end without willing the means [Korsgaard 1997b, 238]. Now this view would be incoherent, if Korsgaard meant that 'wills the means' was literally part of the analysis of 'wills the end', in the way that 'unmarried' is part of the analysis of 'bachelor'. Since 'unmarried' is

[13] After arguing against the 'folk' belief that Kant is arguing that the concept of willing the necessary means is contained in the concept of willing the end, Hill straightaway concludes that the passage supports the Wide Scope reading. He seems not to have noticed that there is any further issue at stake in the interpretation of the passage, and further seems to think that this point is the only controversial aspect of any part of his interpretation worth mentioning. He says almost as much [Hill 1989, 366].

[14] Compare note 11.

part of the analysis of 'bachelor', it simply isn't possible for there to be a married bachelor. So Korsgaard must mean something else by her claim about conceptual inclusion.

I think that the clue to what she means comes from what she says in the rest of the same sentence: "if you will the end and yet fail to will the means to that end, you are guilty of a practical contradiction" [Korsgaard 1997a, xvii]. So Korsgaard doesn't seem to think that willing the end but not willing the means is like being a married bachelor. She seems to think that it is like *believing* that John is a bachelor but not believing that John is unmarried. This is a contradiction (at least, a failure to believe the consequences of one's beliefs) not because of the analysis of 'believes that John is a bachelor', but because of the analysis of 'John is a bachelor'. So for the analogy to be complete, Korsgaard must apparently be thinking of the analysis as not literally of 'wills the end', but of some kind of distinctive practical *content* of willing an end. And this interpretation of Korsgaard is well-supported, I think, by her 'solution' in Korsgaard [1997b, 245–247] and by her statement of her positive view in Korsgaard [1996].

This leaves a serious and intriguing question about what Kant meant by 'analytic' in the practical domain that there is no space here to adequately explore. But I want to consider Kant's specific claim that,

> [w]hoever wills the end, wills (so far as reason has decisive influence on his actions) also the means that are indispensably necessary to his actions and that lie in his power. *This proposition*, as far as willing is concerned, *is analytic*. [4:417, italics added]

The chief problem with Korsgaard's proposed account, as I understand it, is that although it makes *something* turn out to be analytic, it doesn't make *this* proposition turn out to be analytic.

For according to the Korsgaardian conception of analyticity in the practical domain, what is analytic is something like 'if WILL(the end), then WILL(the necessary means)', where 'WILL' expresses the distinctive practical *content* of willing something, of which the account is supposed to provide an analysis. And if this is right, then those who will an end but don't will the necessary means are relevantly analogous to those who believe that John is a bachelor but don't believe that he is unmarried. Now it may be *irrational* to be like this. It may be that you *ought* not to be like this. But it isn't *analytic* that it is irrational to believe that John is a bachelor but not that he is unmarried. At least, if it is, we need an analysis of rationality in order to see why. Similarly, even if this Korsgaardian account is correct, it simply isn't *analytic* that it is irrational to will the end but not will the means. But Kant clearly claims that it is.

I find the alternative conception of analyticity in the practical domain that we seem to need to attribute to Korsgaard greatly intriguing. Unfortunately there is no space, here, to adequately address its merits. But for the main reason just cited, I don't see how it would be able to aid us in explaining the *analyticity* of 'If you will the end, then you ought to take the necessary means'. So I conclude that since Kant does claim it to be analytic, he must understand it as taking Consequent Scope, and his argument must work in the straightforward way that I've outlined: he analyzes 'wills the end', and finds 'ought to take the necessary means' as part of the analysis.

9.7 So Far

This, finally, gives us three strong considerations in favour of Consequent Scope over Wide Scope. (1) Only it, I argued, can make sense, rather than foolishness, of Kant's assertion that it is the categoricity of the moral imperative which is the, or even *a*, distinguishing feature which merits it special treatment. (2) Moreover, the distinction between assertoric and problematic hypothetical imperatives only makes sense on the Consequent Scope Reading. And finally, (3) Kant's argument can only be made proper sense of, given his conception of analyticity, as an argument for the analyticity of Consequent Scope. Perhaps other arguments might have shown that Wide Scope is analytic, but he doesn't appear to have given them. The argument analyzes the wrong concept to be an argument for Wide Scope.

So I take it that we would need rather strong reasons to reverse our interpretation in favour of Wide Scope. But so far, we have still to come across *any* reasons to attribute Wide Scope to Kant. I hold that this is not a coincidence. Wide Scope has been attributed to Kant for one principal reason only, and it is not based in the text at all. This consideration, as I'll argue in the next section, is plausibly much less compelling for Kant than for the contemporary theorists who offer it, and my interpretation of the analyticity argument at [4:417] lets us immediately see why. Since he has a way out of the problem which they lack, it's anachronistic to think that the considerations which compel them would also have compelled him.

9.8 Detaching

The problem, as contemporary theorists see it,[15] is that Consequent Scope has commitments which are simply intolerable. Take the case of someone who wills the end that she robs a bank at gunpoint. If we accept Consequent Scope, then all we have to do is to apply *modus ponens* in order to yield the conclusion that she ought to bring a gun. Take the case of someone who wills to be an axe-murderer. If we accept Consequent Scope, then all we have to do is to apply *modus ponens* to yield the conclusions that he ought to sharpen his axe and that he ought to stake out victims. Indeed, that he ought to swing his axe at someone. According to the lingo, deriving particular 'ought' statements from hypothetical imperatives and their antecedent conditions is called *detaching*. The worry is that Consequent Scope lets us detach *too much*. It lets us detach 'ought'-statements which are patently false. So it must be false.

Wide Scope, on the other hand, doesn't have this feature. Nothing follows from the fact that someone wills to be an axe-murderer, and ought to ensure that he either takes the necessary means to being an axe-murderer or doesn't will to be one. It certainly doesn't follow that he ought to take the necessary means. Perhaps the thing for him to

[15] Gensler [1985], Broome [1999], and Chapter 7, this volume, give perspicuous treatments of the motivation which follows. Hill [1973] spells out carefully how these considerations get applied to the interpretation of Kant. Similar considerations can be found in Darwall [1983] and Korsgaard [1997b].

do is to cease willing to be an axe-murderer. That is what the Wide Scoper would like to say. Since Wide Scope lets her say it, and Consequent Scope seems not to, she infers that Wide Scope has got to be the way to go. It lets us maintain that there is an important normative connection between ends and means, but all it does is to *transfer* the force of reasons from ends to means, as Stephen Darwall puts it [Darwall 1983]. Willing to do things one patently ought not to do doesn't make it the case that one ought to do these things.

And Kant, as I read him, would agree wholeheartedly about all of this. *Except* the bit about any of it telling against Consequent Scope. It's only an argument against Consequent Scope, after all, insofar as Consequent Scope is committed to thinking that it's *possible* to will ends to which one ought not to take the necessary means. The contemporary naturalists about which Wide Scopers are usually worried have to think this. They think that it's possible to give an account of having an end in wholly non-normative terms. Willing an end is, for example, desiring it. Or desiring to desire it [Williams 1981; Frankfurt 1971; Lewis 1989]. Or both. Or it's valuing it, where this *sui generis* mental state is explained by its cognitive and evolutionary role [Watson 1975; Gibbard 1990]. Or it's a desire which would survive cognitive psychotherapy or exist in reflective equilibrium [Brandt 1979; Smith 1994]. Whatever the proposal, the naturalist is going to fall short of concluding that having totally immoral or irrational ends is completely impossible, and so—so far as it goes—the objection that he will detach too much is a good objection against him, if he wants to believe in Consequent Scope.

But Kant isn't this kind of naturalist. In fact, as I've already argued in the previous section, the proper way to understand his argument for the possibility of hypothetical imperatives commits him to thinking that one simply *can't* will an end, unless it could be the case that one ought to take the necessary means. So the scenario that the Wide Scoper envisions is one that Kant will simply find impossible.

What makes it impossible, after all, is the *Categorical* Imperative. The Categorical Imperative is, after all, the centerpiece of Kant's ethical theory. It sets forth the *constraints* on what ends an agent can will. If an end doesn't pass the test of the Categorical Imperative, then it can't be *willed*. The whole idea of the Categorical Imperative is that it is derived merely on the basis of constraints that any will would have to satisfy.

Now it's not that there isn't a sense, for Kant, in which you can will bad ends—you certainly can. For the Categorical Imperative is the expression of a *normative* law, and one which its subjects—imperfectly rational wills—do not necessarily obey. Kant uses the term *willkür* to refer to the will conceived of as the generic capacity for choice. In this sense, animals as well as humans have wills. The animal *willkür*, however, is merely set by its desires. If an animal desires something, then that automatically becomes its end. And that condition is *heteronomy*. Heteronomy of the will is the condition of having one's *willkür* set by the object of one's desires. And this contrasts, for Kant, with *autonomy*. Humans have a capacity for autonomy of the will which animals lack, according to Kant, because in addition to mere desires, their *willkür* can also be governed by *incentives* set by their *wille*. The *wille* is the will conceived of as the faculty for

determining the *law* by which you will be guided. And what the Categorical Imperative tells us, is what things could possibly be products of our *wille*, since it tells us what things could possibly be laws, and what the *wille* does is to determine laws. So though bad ends can be set by your *willkür* and thus be the product of your *choice*, they can't be the product of your *wille*. You can't *will* them, in the sense that I claim is appropriate to hypothetical imperatives, on Kant's view.

I am also not claiming that Kant does not think that hypothetical imperatives come into play with respect to bad ends. As Kant himself says, "[t]he prescriptions needed by a doctor in order to make his patient thoroughly healthy and by a poisoner in order to make sure of killing his victim are of equal value so far as each serves to bring about its purpose perfectly" [4:415]. This is true, when these technical hypothetical imperatives are understood as precisely that—*hypothetical*. If hypothetical imperatives are truly analytic, then they don't discriminate between conceptually possible ends, and so it follows that *if* you have the end of poisoning, *then* you ought to take the necessary means. But that doesn't mean that anyone *does* will this end, in the required sense. Indeed, on my reading, nothing in Kant's analytic argument for the possibility of hypothetical imperatives shows that anyone is under *any* assertoric hypothetical imperative. For he needs a transcendental argument for our autonomy, in order to establish that we are even the kinds of being to *have* ends—i.e., in his view, to be capable of guiding ourselves in accordance with the *categorical* imperative—in the first place.

The claim that I'm trying to make in this section is small, and shouldn't be understood as resting on any particular interpretive hypotheses about Kant's various accounts of the Categorical Imperative, his account of autonomy, or his moral psychology. I've simply pointed out that it follows from my interpretive claim about the argument at [4:417] of the *Groundwork* that the kinds of disturbing cases that give rise to the worry about detaching too much *can't* arise, for Kant. That argument stands or falls by itself. Here I'm only trying to place the resulting view in the context of Kant's larger project, and to explain why this is not a *crazy* thing to understand Kant as thinking. The point, after all, is a defensive one.

I don't claim that this is the only way of understanding the relationship between hypothetical imperatives and the Categorical Imperative, for Kant. But I do claim that it is enough to show that if he likes, Kant *can* get his solution to the 'detaching' problem for free. His moral theory is already deeply engaged in the project of setting constraints on what can be the product of an autonomous will, and his moral psychology specifically locates in the human will a part that is the source for such autonomous willing. And if my interpretation of the analyticity argument at [4:417] is right, then this is what Kant *has* to say. For it simply follows from that interpretation that you can't will ends unless you ought to take the necessary means to them. So 'detaching' too much turns out not to be a problem for Kant at all. It's certainly not one worth overriding all of the serious textual evidence *against* thinking that Kant's hypothetical imperatives are Wide Scope.

9.9 A Final Thought

Korsgaard and Hill, among others, have attributed to Kant a special Principle of Practical Reason—the Hypothetical Imperative, which governs the instrumental realm in the way that the Categorical Imperative governs the moral realm. This may or may not be slightly suspicious, given that Kant tells us in the second *Critique* that the Categorical Imperative alone is "the fundamental law of pure practical reason" [5:31]. But either way, the foregoing suggests that there is a legitimate sense in which making such claims might be a mistake. Whether it is, depends on what we take Kant to mean by 'principle'. We might think that principles of practical reason are simply truths about practical reason, and their being objective is a matter of their being properly judgeable by any rational being. If this is what it is to be an objective principle, then Wide Scope fits the bill. It is a truth about what people ought to do, and since it is analytic, anyone possessing the concept of willing an end must rationally accept it.

But it often seems that Kant, and his interpreters, mean something quite different by principle. Korsgaard, for example, tells us that "[t]he familiar view that the instrumental principle is the *only* requirement of practical reason is incoherent" [Korsgaard 1997b, 220]. She means to be arguing, in part using Kant as an authority, that it is incoherent to think that there are only hypothetical imperatives, as the naturalists discussed in the last section seem to believe. The 'instrumental principle' is supposed to be the Hypothetical Imperative. But here she is claiming that the instrumental principle is a *requirement*. This seems to presuppose a much different sense of 'principle' than the innocuous sense of the last paragraph. Consequent Scope is not a requirement. It may be a truth about requirements, and about when agents are under them. It says, after all, that if an agent wills an end, then she is required to take the necessary means. But it is not *itself* the requirement that the agent take the necessary means, just the fact that there is such a requirement.

Kant also sometimes seems to have this sense of 'principle' in mind. Sometimes he seems to use 'principle' merely in such a way as to overcome the generality problem generated by the fact that imperatives apply only to *finite* rational beings, and not to infinite rational beings who always do what they ought. Principles, in this sense, are like we have been understanding imperatives to be. They *apply* to agents, *directing* them to do one thing or another. Although Consequent Scope is *about* agents, however—it says when and how they can come under obligations to do one or another thing—it does not itself *apply* to them. It does not *direct* them to do anything, in the way that Wide Scope claims that there is some directive applying to any agent, directing her to either take the means to her ends or else give them up. If this is the sense of 'principle' that interpreters like Korsgaard and Hill have had in mind, then I've been arguing that there is no fundamental principle of instrumental reason, no Hypothetical Imperative, capital-H capital-I.

The moral, then, if there is one, is that it's no coincidence that Hill has to stretch to come up with a passage to cite in which Kant refers to 'the' hypothetical imperative, in the singular.[16] It's not *the* Hypothetical Imperative, binding on everyone, but *hypothetical imperatives*, binding on those with certain ends, which interest Kant. Or so I've tried to illustrate.[17]

[16] He does stretch. He cites only one passage, and cites it twice, but both times removes it from context which makes it look like Kant is really referring to more than one imperative in the quoted passage.

[17] Special thanks to Maurice Goldsmith, David Sussman, Stephen Darwall, Ant Eagle, Gideon Rosen, graduate student audiences at Princeton University and at the University of California at Berkeley, and to two anonymous reviewers for the *Australasian Journal of Philosophy*.

10

Hypothetical Imperatives, Scope, and Jurisdiction

10.1 Hypothetical Imperatives vs *the* Hypothetical Imperative

The last few decades have given rise to the study of *practical reason* as a legitimate subfield of philosophy in its own right, concerned with the nature of practical rationality, its relationship to theoretical rationality, and the explanatory relationship between reasons, rationality, and agency in general. Among the most central of the topics whose blossoming study has shaped this field, is the nature and structure of *instrumental* rationality, the topic to which Kant has to date made perhaps the largest contribution, under the heading of his treatment of *hypothetical imperatives*.

After forty years, Tom Hill's 1973 article 'The Hypothetical Imperative' remains one of the best entrees into the issues surrounding instrumental rationality, as well as the main voice of probably the most mainstream interpretation of Kant's own view of hypothetical imperatives.[1] Hill's article offered not only an interpretation of Kant's theory of hypothetical imperatives, but a general account of why, in Kant's view, demonstrating the possibility of a categorical imperative requires a synthetic argument of the kind that he proceeds to give in section 10.3 of the *Groundwork* and in the *Critique of Practical Reason*—showing that understanding Kant's view of hypothetical imperatives is essential for understanding the structural issues in his practical philosophy more generally.

The main theme of Hill's article is that just as Kant's moral philosophy relies on an important distinction between categorical imperatives and the Categorical Imperative, Kant's broader practical philosophy also relies on an equivalent distinction between hypothetical imperatives and the Hypothetical Imperative (hence the title).[2] For Hill's

[1] My own engagement with Hill's article actually led to both of my first two publications in philosophy, including both my not-so-creatively titled dissent, 'The Hypothetical Imperative?' (Chapter 9, this volume), and 'The Scope of Instrumental Reason' (Chapter 7, this volume), whose origin was as a splinter off of the other paper.

[2] Curiously, Kant actually never unambiguously refers to 'the' hypothetical imperative, in the singular, as he does to the Categorical Imperative. Hill claims otherwise, twice citing a single passage in which the term

view it is important not only that there is something called 'the Hypothetical Imperative', but that it is itself an imperative. Whereas individual hypothetical imperatives, such as 'if you want to lose weight, then count your calories' and 'if your aim is to graduate, then you ought to attend your classes' enjoin particular, concrete, actions to agents with particular ends, according to Hill the Hypothetical Imperative says simply: 'take the necessary means to your ends'. Rather than enjoining particular concrete ends, it enjoins simply taking the necessary means to your ends, whatever those are, and rather than addressing only the people with one end or another, it addresses every rational agent as such.[3]

It will be helpful to compare imperatives, which after all for Kant are simply the expression of laws to imperfect wills which do not necessarily obey them, to more mundane laws. One important feature of laws is that they have *jurisdictions*. For example, in the state of New York, it is illegal to turn right at a red light. The jurisdiction of that law is *drivers in New York*, and what it prohibits is *turning right on red*. In general, anyone who is simultaneously a driver in New York and is turning right on red is in violation of this law, but it is important to appreciate that being a driver in New York and turning right on red make different contributions to this fact. If you are a driver in New York and you *don't* turn right on red, then you are *complying* with the law, whereas if you are a pedestrian in New York or a driver in Buenos Aires or Cairo, the law simply doesn't apply to you. The reason why drivers in Cairo who turn right on red aren't in violation of New York traffic laws is that the New York state legislature doesn't have *jurisdiction* over drivers in Cairo—not that it does have jurisdiction, but they are in compliance.

According to Hill, the significance of the Hypothetical Imperative (in the singular, with capitals) is that its jurisdiction is just as universal as the Categorical Imperative—it has jurisdiction over absolutely all rational agents, no matter what they are like. The reason this shapes Hill's broader interpretation of Kant is that this makes the Hypothetical Imperative sound an awful lot like Kant's description of categorical imperatives—which are universal laws, with jurisdiction over every rational agent, no matter what she is like. For on one highly eligible interpretation of Kant it is precisely the universal jurisdiction of categorical imperatives which makes them so hard to argue for and explain, but Hill cannot accept this interpretation, for if it were right, then hypothetical imperatives would be just as hard to argue for and explain, since they derive from the Hypothetical Imperative, which shares with categorical imperatives the feature of having universal jurisdiction.[4]

'the hypothetical imperative' appears; unfortunately, the passage is taken out of context, and comes from the second half of a long German sentence whose first half refers alternately to 'the imperative of prudence' and 'the technical imperative', and in which the most natural reading of 'the hypothetical imperative' in the second half is as a *bound* reading, so that it refers alternately to either the imperative of prudence or the technical imperative. Nothing much should hang on this issue.

[3] Compare also pp. 51–57 of Hill and Zweig [2002], where this view is reiterated.

[4] Compare Hampton [1998, 165–166], who argues on just these grounds that Kant's view that hypothetical imperatives are easier to explain than categorical imperatives is incoherent (boldface added for emphasis):

> Kant's position on the nature of hypothetical imperatives must be construed (**contra his explicit wishes**) such that understanding the bindingness of a hypothetical imperative is no

In earlier work I argued that Hill's textual interpretation of Kant's account of hypothetical imperatives is under-motivated and fails to account for a number of interesting facts about the text, including the details of Kant's analytic argument for hypothetical imperatives, the significance of his distinction between *problematic* and *assertoric* hypothetical imperatives, and changes in his treatment over time, particularly leading up to the publication of the *Critique of Judgment*.[5] My goal here is not to rehash those arguments, but to explain why it *is* so important—not only for Kant, but for contemporary efforts to explain requirements either of rationality or of morality—what we take the underlying *jurisdiction* of those requirements to be. My main focus will be on the constraints on *any* theory that are placed by the answer we give to this question, rather than on interpretive questions about Kant, but I will try to indicate along the way why it is fruitful to understand Kant as having been motivated by precisely the issues that I will be trying to make vivid, here.

The remainder of the paper consists of five sections; in section 10.2 I'll introduce a mostly familiar dialectic framed in contemporary terms about the structure of hypothetical imperatives, framed in terms of their *scope*, and try to get clearer about how they are related to one another. One of the most important points to be made in that section is that in order for questions about scope to be interesting, it helps a great deal to impose certain constraints on the proper interpretation of the normative concept that we use to formulate them. In section 10.3 I'll argue that the law gives us a model for the kind of normative concept that allows the interesting scope question to be asked, and that this is because it makes room for non-vacuous distinctions in *jurisdiction*.

Then in section 10.4 I'll discuss the relationship of *scope* to *jurisdiction*, and lay out two broadly Kantian concerns one might have—and which I think we should have—about how to explain Wide scope rational requirements with universal jurisdiction. In section 10.5 I'll offer a positive, Kant-inspired but not necessarily strictly Kantian picture which, by giving one interpretation of what makes Narrow scope statements of conditional requirements of rationality true, gives us a rich and satisfying alternative to the concerns raised in section 10.4. Finally, in section 10.6 I'll close by revisiting my original disagreement with Hill, and suggesting that in some (particularly important!) respects, our readings of Kant may not actually be so far apart.

> easier than understanding the bindingness of a categorical imperative. My interpretation cannot save Kant's belief that the former is more straightforward than the latter; **indeed, my argument is that Kant's belief is wrong**. The only way to analyze Kant's analyticity claim is to do so in a way that locates in hypothetical imperatives the same mysterious objectivity that attends the categorical imperative. Even more strikingly, I have argued that the force of hypothetical imperatives is dependent on, and *is at least in part constituted by*, the force of some antecedent categorical imperative that is in part definitive of instrumental rationality.

[5] Chapter 9, this volume.

10.2 Instrumental Rationality: Some Possible Views

The topic of instrumental rationality explores how what we ought to do depends on our ends. It is actually plausible that there are in fact several different closely related topics which can loosely be described in these terms and have sometimes been confused,[6] but glossing over complications, we may take the basic datum to be that at least in paradigmatic cases, there is something going wrong with someone who intends some end, believes that some means are necessary for that end, and has no intention whatsoever for the means.

To simplify things by avoiding quantifying over ends and means, I will assume that *e* and *m* are arbitrary actions, that E*x* is the proposition that *x* intends to do *e*, M*x* the proposition that *x* intends to do *m*, and B*x* the proposition that *x* believes that *m* is a necessary means to *e*. Also for simplicity, and glossing over several complications, I will follow a common convention in deontic logic, and assume that we can represent the claim that *x* ought to do *a* as OUGHT_x(*x* does *a*), treating 'OUGHT_x' as a propositional operator.[7] These simplifications make it possible to distinguish clearly between four important kinds of view from the literature about instrumental rationality, distinguished from one another by the *scope* of the 'ought':

Narrow: $\sim\forall x\,(Ex\&Bx \rightarrow \text{OUGHT}_x(Mx))$
Intermediate: $\sim\forall x\,(Bx \rightarrow \text{OUGHT}_x(\sim Ex \vee Mx))$
Wide: $\sim\forall x\,(\text{OUGHT}_x(\sim Ex \vee \sim Bx \vee Mx))$
Myth: $\sim\forall x\,(\text{OUGHT}_x(\sim Ex) \vee \text{OUGHT}_x(\sim Bx) \vee \text{OUGHT}_x(Mx))$

Controlling for differences of opinion about exactly how 'OUGHT_x' is to be understood, I've advocated a view with the structure of Narrow in Chapter 8, this volume, Jonathan Way has advocated an attractive view with the structure of Intermediate in Way [2010], Joseph Raz [2005], and Niko Kolodny [2008a], [2008b] have advocated views something along the lines of Myth, and Wide has been widely endorsed, including by Hill [1973], Darwall [1983], and Broome [1999], among other prominent proponents.

Proponents of Wide sometimes say that only Wide is uncontroversial, but that depends on how we interpret 'OUGHT_x'. If we interpret 'OUGHT_x' as meaning '*x* won't do everything that *x* rationally ought to do unless', then Wide is actually *entailed* by each of the other views. After all, if *x* won't do everything that *x* rationally ought to do unless ~E*x*, it is clear that *x* won't do everything that *x* rationally ought to do unless ~E*x* ∨ ~B*x* ∨ M*x*. So on this reading, the first disjunct of Myth entails Wide, and similar reasoning goes for the other disjuncts. Similarly, if the truth of E*x*&B*x* guarantees that *x* won't do everything that *x* ought unless M*x*, *x* certainly won't do everything that *x* ought

[6] Including by me; see Chapter 7, this volume.

[7] I don't believe this simplification affects anything important in this paper, but for some of the reasons why it is arguably an oversimplification, see Schroeder [2011].

unless either M*x* or it is not the case that E*x*&B*x*—so on this reading, Narrow also entails Wide—and similar reasoning goes for Intermediate. So if we interpret 'OUGHT_x' in this way, Wide *isn't* controversial; it is just a description of the data that everyone wants to be able to explain.[8]

Now, the problem is not that we cannot formulate *some* dispute between Wide and Narrow scope views about instrumental rationality, by use of a normative concept like '*x* will not do everything that *x* rationally ought to do unless'. Using such a concept, as I've noted, all sides agree that Wide is true. But there can still be a very real debate about whether Wide is the *only* true thing to be said, at this level of generality. However, prominent objectors to the Wide scope view in the literature—including Niko Kolodny, Joseph Raz, and myself—have not presented their view as granting that Wide is true but insisting that one of the other theses is also true. On the contrary, many remarks of these critics suggest that they have been trying to argue directly against Wide. So it seems that charity requires taking them to interpret 'OUGHT_x' as meaning something else.

And indeed, though Wide is uncontroversial when we interpret 'OUGHT_x' in this way, what *is* controversial is what *makes* this true. On this interpretation of Wide, for all that it says, the explanation of why *x* will not be doing everything that *x* ought unless ~E*x* ∨ ~B*x* ∨ M*x* works the same way as the explanation of why someone who ought to post a letter will not be doing everything that she ought unless she either posts it or burns it.[9] But if someone who has promised to post a letter (and hence ought to post it) burns it instead, she is either posting it or burning it, but there is nothing intuitive about the claim that she is satisfying one of her obligations.[10] Moreover, if she does post the letter, there is intuitively nothing *extra* that she does right, in virtue of the fact that she thereby either posts it or burns it. So when proponents of Wide say things like that you can satisfy the requirements of instrumental rationality just as well by giving up your ends as by going on to intend the means,[11] or that there is something *distinctive* going wrong with someone who fails to intend the believed means to her intended end,[12] it does not make sense to interpret them as meaning simply that *x* will not do everything that *x* rationally ought unless ~E*x* ∨ ~B*x* ∨ M*x*—they must mean something stronger.[13] It is therefore this stronger thing that detractors of Wide scope views, including Joseph Raz, Niko Kolody, and myself have meant to deny.[14]

[8] The data is actually more complicated; I'm simplifying by ignoring what happens when, for example, you believe that the means is necessary for the end but that you will do it without intending to do so. See Setiya [2007a] and Kolodny [2008a] for more discussion of these sorts of important details, over which I will proceed to gloss.

[9] This is Ross's Paradox, originally raised in the context of imperatives. See Ross [1941].

[10] Although compare Wedgwood [2007], chapter 4, on how easy it may be to satisfy *some* obligation versus how hard it is to satisfy *all* obligations.

[11] Compare Hill [1973], Broome [1999].

[12] Compare Wallace [2001].

[13] For this point in connection with Wide, see van Roojen [unpublished]. Similar issues were originally introduced in the context of understanding conditional 'oughts' as oughts with material conditional complements within Standard Deontic Logic by Chisholm [1963].

[14] See particularly Raz [2005], Kolodny [2008a], [2008b], and Chapters 7 and 8, this volume.

10.3 Scope in the Law

This raises the question of just what this stronger interpretation of 'OUGHT_x' is supposed to be. And for the answer to that, it is helpful to look back to the example of the law. Just as someone who intends *e*, believes that *m* is a necessary means to *e*, and does not intend *m* is not entirely as she rationally ought to be, someone who is a driver in the state of New York and turns right on red violates the New York state traffic regulations. If we let E*x* be the proposition that *x* is driving, B*x* be the proposition that *x* is in the state of New York, and M*x* be the proposition that *x* does not turn right on red, we can now interpret Narrow, Intermediate, Wide, and Myth as accounts of what is going on in this case. Again, if we interpret 'OUGHT_x' as meaning '*x* will be in violation of New York state traffic regulations unless', there is nothing controversial about Wide. But again, since it is also true that you will be in violation of New York state traffic regulations unless either you are not in New York or you are not a driver or you do not turn right on red or you drive a convertible, this does not seem like a particularly interesting thing to say.

On the other hand, if we interpret 'OUGHT_x' as 'New York state traffic regulations require *x* to', then Wide is not obviously true after all. On the contrary, Narrow and Intermediate are much more natural views. According to Narrow, New York state traffic regulations have jurisdiction over drivers in New York, and require them not to turn right on red. According to Intermediate, New York state traffic regulations have jurisdiction over everyone in New York, driver or not, and require them to not turn right on red while driving. Both of these views are plausible, because it is (relatively) easy to understand why the New York legislature has jurisdiction over people in the state of New York—particularly if they are driving. But on this interpretation, Wide says that New York state traffic regulations have jurisdiction over drivers in Buenos Aires and Cairo, and require them to either not turn right on red or else not be in New York. This claim is particularly implausible, because it is very hard to see how the New York state legislature could have gotten jurisdiction over drivers in Cairo or Buenos Aires. The much more natural way of understanding what is going on in this case is therefore that Wide is not, in fact, true on this interpretation.

Another way of seeing the same thing, I think, is to observe that drivers in Cairo and Buenos Aires are not *complying* with New York state traffic regulations, simply because they are not in New York. In contrast, drivers in New York who don't turn right on red *are* complying with New York state traffic regulations. It is true that there are two ways to *avoid violating* New York traffic regulations—you can refrain from turning right on red, or you can leave the state. But these are not two ways of *complying* with the regulations. One is compliance, and the other is escape. The distinction between compliance and escape tracks the regulations' *jurisdiction*, because you can comply with a regulation only if you fall under its jurisdiction, and leaving the jurisdiction of the regulation is sufficient to avoid violating it.

What these remarks illustrate, I believe, is that the concept of what is required by New York traffic regulations is the right kind of concept to allow for a meaningful

scope debate precisely because it allows for a non-vacuous distinction between who falls inside and outside the jurisdiction of the law. Any concept that allows, at least in principle, for a non-vacuous jurisdiction distinction enables us to have a real scope debate, because the concept of jurisdiction guarantees that no one who falls outside the jurisdiction is bound by any of its requirements. And that means that all requirements are conditional on falling under the relevant jurisdiction. So if proponents of the Wide scope view are willing to use such a law-like concept, and hold that something—a kind of means–end consistency—is required of *all* rational agents, then they are committed to a view about the jurisdiction of this requirement or its source—that it has jurisdiction over all rational agents.

What I've just been arguing, is that the dispute between advocates of Wide and its detractors is not best understood as a dispute about whether Wide is true, if interpreted in such a weak way that it follows from Narrow, Intermediate, or Myth, and moreover that understanding this dispute requires drawing an important distinction between *compliance* and *avoidance*, which tracks the important concept of *jurisdiction*. The distinction between compliance and avoidance is precisely what makes the dispute among proponents of Narrow, Intermediate, Wide, and Myth an interesting one.

10.4 The Impact of Jurisdiction on Explanatory Resources

So far I've been arguing that insofar as there is an intelligible debate about the scope of instrumental rationality, it is a debate about the scope of a concept that is *law*-like, in that it distinguishes between violation and non-compliance, and that this distinction is closely related to the concept of *jurisdiction*. The fact that law-like normative concepts are precisely of the right kind to allow for these important distinctions should encourage us, I believe, about whether these distinctions are important for Kant's own theory of instrumental rationality. For Kant himself employed a rich terminology deeply indebted to thinking in terms of the concept of law. Imperatives are defined outright in the *Groundwork* as the expression of laws to imperfectly rational wills.

The reason why scope is closely related to jurisdiction is that a law only needs to have jurisdiction over those who satisfy its condition. Since Narrow and Intermediate only postulate laws given some substantive condition, the laws they postulate need only have narrow jurisdiction. But since Wide postulates an unconditional law, that law must have universal jurisdiction. This is precisely what makes the Wide interpretation of New York state traffic laws so implausible—for it is implausible that the New York state legislature has the authority to legislate laws with jurisdiction over drivers in Cairo or Buenos Aires.

So what about the Wide interpretation of the requirement of instrumental rationality? Is it plausible that it has universal jurisdiction over all rational agents? I have to confess that I find this idea as puzzling as I do the idea that the New York state

legislature has jurisdiction over drivers in Buenos Aires. What is the source of this universal requirement supposed to be, and how does it acquire jurisdiction over every rational agent? I find it difficult to even get my head around this question.

Moreover, I think that this sort of puzzlement is distinctly Kantian. For according to Kant, rational agents are autonomous, in the sense that they act according to laws *that they set for themselves*:

> The will is therefore not merely subject to the law, but subject in such a way that it must be considered as also *giving the law to itself* and **only for this reason** as first of all subject to the law (of which it can regard itself as the author).[15]

But if hypothetical imperatives derive from a master Hypothetical Imperative with universal jurisdiction over every rational agent, then their authority seems to come from outside the agent—for it comes from whatever authority has jurisdiction over *all* rational agents. So although it is possible that this is based on a misconception on my part, such an external source of rational requirements sounds *prima facie* exactly like *heteronomy* of the will. Autonomy, in contrast, would be each agent being bound only by laws she sets for herself—that is, the idea that each agent falls only under her own jurisdiction.

A second, related but also important question is how bare rational agency suffices to explain why someone falls under this particular requirement. Kant explains in the *Groundwork* that the hypotheticality of hypothetical imperatives is exactly what makes them unpuzzling, and the kind of thing whose possibility can be established by analytic means. The 'objective necessity' which they present, he says, is only 'based on a presupposition'—a presupposition that is rich enough for us to know that someone who satisfies it is indeed bound by that imperative.

> By contrast, 'How is the imperative of morality possible?' is beyond all doubt the one question in need of a solution. For the moral imperative is in no way hypothetical, and consequently the objective necessity, which it affirms, cannot be supported by any presupposition, as was the case with hypothetical imperatives.[16]

I believe that the insight behind this passage is that the stronger the condition on which an imperative applies, the richer the explanatory materials that we will have, in order to explain why and how it is, that the imperative applies. What makes categorical imperatives puzzling and in need of explanation, on this view, is precisely that they are unconditioned, and so we have nothing more to work with than rational agency as such, in seeking to explain them. So Hill's Hypothetical Imperative, which unconditionally requires means–end coherence of every rational agent, should be puzzling for exactly the same reason, and I think that it is.

[15] Kant [2002, 232] (4:431). Boldface added for emphasis; italics in original.
[16] Kant [2002, 220] (4:419).

So I believe that there are two, distinct but closely related, puzzles about unconditional requirements like that postulated by the Wide scope view of instrumental rationality, as I've been interpreting it in this paper. The first is the puzzle of what could have jurisdiction over every rational agent, and the second is how the bare idea of rational agency as such could suffice to explain not only why an agent falls under this jurisdiction, but why the requirement is in force for her. Since I believe that these puzzles are closely related, however, I don't think it should be surprising if a single move addresses both.

10.5 Autonomy of the Will

Whereas it is puzzling, I think, how something could come to have jurisdiction over every rational agent, I don't think it is similarly puzzling how a rational agent could come to have jurisdiction over herself. This is not to say that there are *no* philosophical puzzles about the latter—on the contrary, if this is Kant's view, that I have authority over myself is something that I may only be able to establish through a transcendental argument—but only that the same puzzles do not arise. If an agent has jurisdiction over herself, then she can create rules or laws for herself. In the helpful terminology of Sam Shpall [2013], [forthcoming], she can rationally *commit* herself to acting in one way or another.[17]

Because the New York state legislature has jurisdiction over drivers in the state of New York, it has the authority to require things of them. But in order to exercise this authority, it must act. If it passes legislation which says, 'drivers may not turn right on red', then drivers in New York become required not to turn right on red. And if it later passes legislation which says, 'drivers may turn right on red', then this former requirement is relaxed. Similarly, if a rational agent has jurisdiction over herself, then she has the authority to create rules for herself. But in order to exercise that authority, she must do something. So in order for an agent to have meaningful such authority over herself, there must be something that she can do to exercise it.

The virtue of the Narrow and Intermediate accounts of instrumental rationality is that they tell us exactly what a rational agent must do, in order to exert this authority over herself. On the Narrow view, what she must do to commit herself to intending *m*, is to intend *e* and believe that *m* is necessary for *e*. On the Intermediate view, what she must do to commit herself to not both intending *e* and not intending *m*, is to believe that *m* is necessary for *e*. Because each of these views makes the requirement governing an agent not only conditional, but conditional on something that agent does (in a very expansive sense of 'does' which includes belief and intention), they are very naturally construed as simply telling us what an agent must do, in order to commit herself.

[17] Shpall [2013] argues convincingly for the importance of our intuitive notion of commitment in considering cases like instrumental rationality and enkrasia, and in [forthcoming], he develops a rich characterization of the distinctive and important features of this sense of commitment.

This picture does away with the idea that there is some peculiar source of rational requirements which somehow has jurisdiction over every rational agent, and replaces it with the idea that each rational agent has jurisdiction over herself. In that way, it addresses the first problem which puzzled me in the last section. And it makes good on Kant's observation that was at the heart of the second problem from that section. Kant's observation, I suggested, was that conditional requirements give us richer explanatory resources in order to explain requirements, and on the picture being developed in this section, we utilize those explanatory resources by conceiving of them as the things an agent must do, in order to exercise her authority over herself.

I don't know that Kant's own view is anything like this, but if it is, I suspect that it has neither the shape of Narrow nor that of the Intermediate view I've described, but rather that of an alternative, Kantian Intermediate view:

Kantian Intermediate: $\forall x\,(Ex \rightarrow \text{OUGHT}_x(\sim Bx \vee Mx))$

After all, if rational agents give themselves laws, on Kant's picture it will not be their beliefs which do so, but the exercises of their will. In willing an end, I commit myself to taking the means I believe to be necessary to it. In ceasing to will this end, I relieve myself of that commitment. So long as I am under the commitment and fail to intend some means I believe to be necessary, I am failing to do everything to which I am committed, and hence am irrational. But that is not because there is any more general rule of rationality which explains why this is so. It is just because through my will, I have the power to rationally commit myself, and rationality is nothing more than living up to my own commitments—that is, being successfully governed by the laws that I set for myself.[18]

10.6 Kantsequences

In this paper I've been trying to argue that it helps to use a law-like normative concept—one which allows for meaningful distinctions about jurisdiction—in order to get to the heart of what has been at stake in the so-called scope dispute over instrumental rationality. And if this is right, then we should not be surprised, given Kant's own reliance on the concept of law, if this dispute is important for his purposes. And I've promoted two Kantian ideas as having real import for this dispute: first, that rational agents are autonomous, in the literal sense that they give themselves their own laws, and second, that the very thing that makes hypothetical imperatives easier to explain is their conditionality, since that provides us with greater explanatory materials. When we put these two Kantian ideas together, we get the view that instrumental rationality is not something that is required of us, as rational agents, but rather, simply a reflection

[18] This description matches the interpretive claims about Kant advanced in Chapter 9, this volume. Compare especially the discussion of autonomy in Hill [1985], and in Hill [1989], especially pp. 140–141.

of our capacity to require things of ourselves. It is not right—or at least not illuminating—to think of instrumental rationality, on this picture, as backed up by a law-like Hypothetical Imperative.

I have not intended the main claims of this paper to amount to interpretive claims about Kant; merely to emphasize what I take to be Kantian themes that I think are independently important in this domain. But if Kant really did accept something like the Kantian Intermediate view, holding that the will is our capacity to require things of ourselves, he certainly did not think that it gives us the authority to require just *anything* of ourselves. Just as the New York state legislature has the authority to require drivers not to turn right on red but lacks the authority to require voters to pay a poll tax, rational agents, though they have the authority to require *some* things of themselves, lack the authority to require themselves to lie, for example, or to neglect their self-development. Famously, for Kant all limits on what agents have the authority to require of themselves must come from the bare condition that all requirements must have the form of a law. So whereas Kant's account of hypothetical imperatives, on this view, reflects the authority that rational agents have over themselves, his account of categorical imperatives reflects the limits of that authority.

And this means, I think, that there is something worth calling the Hypothetical Imperative after all—even though I don't think it is right to think of it as an imperative. After all, at the end of 'The Hypothetical Imperative', Hill emphasizes that what is most important about the claims that he makes in the article is that it shows how the Hypothetical Imperative and the Categorical Imperative are simply two different reflections of a single capacity for practical reason, which can be 'paradoxically' summarized with the edict, 'Do what you will'. But this is exactly what I have just been arguing follows, if anything like the picture of this paper is accurate about Kant's own views. If practical reason is auto-nomous in the way I've described, then the Hypothetical Imperative is a reflection of the *auto*, and the Categorical Imperative is a reflection of the *nomous*. If that is true, it is an important truth, and no paradox.[19]

[19] Special thanks to Tom Hill, Robert Johnson, Mark Timmons, Errol Lord, Andrew Sepielli, and an anonymous referee.

11

Scope for Rational Autonomy

Sometimes we act, intend, or believe irrationally. When we do, it is natural to describe us as having broken the rules of rationality. And violating the rules of rationality, it is widely believed, is some kind of particularly serious failing. But what are these 'rules of rationality'—where do they come from, and from whence do they derive their peculiar force? My aim in this paper is to explore a picture on which the 'rules' of rationality are rules that we lay down for ourselves. Each of us is the author of the rules of rationality that apply to us, and so when we are irrational, the rules that we are breaking are our own. This picture answers both where these rules come from and from whence they derive their force. And the specific version of the picture that I will develop also explains why even though the rules of rationality are up to us, it appears that they are not.

My path to this picture, however, will be somewhat indirect. I will begin with an old dispute about the *structure* of the rules of rationality—the debate about whether such rules have 'wide' or 'narrow' scope. I'll explain and endorse some of the key elements of John Broome's account, in his paper 'Requirements', of what the interesting wide/narrow dispute has all along been about, and use that discussion to pose Broome's two central challenges, in that paper, for narrow scope theories. My diagnosis of the force of Broome's challenges will turn on the fact that he gives us no way to think about the *source* of what he calls 'source requirements'. These challenges look like good challenges because they fail to distinguish between two different kinds of narrow-scope views—ones which differ on the *authorship* of the rules of rationality. After developing a simple model that allows for this important distinction, I'll close by showing how it addresses Broome's worries about narrow-scope rules and provides an answer to the puzzle about why, if we are the authors of the rules of rationality that apply to ourselves, it nevertheless seems that we have so little control over what is or is not rational.

11.1 A Structural Question About the Rules of Rationality

Paradigm instances of irrationality include cases like believing outright contradictory propositions, failing to accept the obvious consequences of your beliefs, failing to

intend to do what you believe you ought, having contradictory intentions, or failing to intend the means you believe to be necessary for your intended end.[1] In each of these cases, you seem to be flouting one or another of the rules of rationality. A familiar question concerns the structure of these rules: are they 'wide-scope' or 'narrow-scope'?[2]

According to the wide-scope answer, the rules of rationality are universal in their domain, applying to everyone equally, but ask only non-specific things of us: to abstain from contradictory beliefs, for example, or to maintain a combination of intentions and beliefs which together exhibit means–end coherence. According to the narrow-scope answer, the rules of rationality are specific in their domain, applying only to people with certain beliefs, intentions, or combinations of beliefs and intentions. But of the people to whom those rules apply, they ask specific things: to have or lack a particular belief, for example, or to have or lack a particular intention.

These views are so-named because if we think of the rules of rationality as having a conditional structure, the views differ over the relationship between the rules and the conditional. According to wide-scope views, what is required by the rule has scope over the conditional, and so satisfying a (material) conditional is thought to be what is required by the rule, but the rule itself is thought to be unconditional—and hence to apply equally to everyone. Whereas according to narrow-scope views, what is required by the rule has narrow scope within the conditional's consequent, and so satisfying the rule requires a particular belief or intention, but the rule itself is conditional in structure, so that it does not apply to someone who does not satisfy its condition. The cases which are most easily thought of in this way include the phenomena of means–end coherence, closure, and non-*akrasia*, which are all often formulated in conditional terms. (For example, 'If you will the end, you are required to will the means you believe to be necessary to that end.') But consistency of belief and intention also fit this model. According to the wide-scoper, there is a rule mandating the property of being such that if you believe p, then you don't believe $\sim p$, whereas according to the narrow-scoper, there is only a rule requiring believing $\sim p$, but that rule only applies to people who believe p. Similar points go for consistency of intention.

It is important, I think, to be clear that there is an interesting debate here quite independently of whether this debate can be expressed using the word 'ought' or the word 'required'. It may be that in whatever sense you count as breaking one of the rules of rationality whenever you have inconsistent beliefs or fail to believe the obvious consequences of your other beliefs, we cannot express these rules of rationality using the words 'ought' or 'required' at all. The debate also cannot be about whether there are any true wide-scope principles. For everyone can agree that the concept, 'you do not obey all of the rules of rationality that apply to you unless' can be used to formulate

[1] Here I'll ignore a number of complications about whether each of these claims needs to be qualified in important ways.

[2] Compare Broome [1999] for a classic discussion, though one that omits important nuances to follow, and Chapter 7, this volume, for a contrary, similarly unnuanced, perspective.

true wide-scope principles.[3] So the interesting wide/narrow dispute is not about how to interpret sentences of some kind, but about the underlying explanation of whatever is true, in this domain.

11.2 Broome on Source Requirements

In 'Requirements', his contribution to the festschrift for Wlodek Rabinowicz, John Broome makes this same point by distinguishing between what he calls 'property' and 'source' requirements.[4] In the 'property' sense, according to Broome, we may say things like 'prudence requires you to save for your retirement', and all that we mean is that necessarily, you are not prudent unless you save for your retirement. It's a familiar fact, sometimes known as the Kanger-Anderson Reduction, that any operator with the structure 'necessarily, you aren't F unless' obeys standard deontic logic. So on the property interpretation, 'rationality requires you to have consistent beliefs' only says that necessarily, you are not rational unless you have consistent beliefs.

This, however, just amounts to a description of the facts about when you are irrational. It doesn't tell us anything about what makes you irrational, when you are. So Broome thinks that 'requirement' also has a *source* sense. In the source sense, 'rationality requires you to have consistent beliefs' means not only that necessarily, you are irrational unless you have consistent beliefs, but it tells us why this is so: because one of the *source* requirements of rationality forbids having inconsistent beliefs. According to Broome, to have the property of rationality that figures in the *property* sense of what 'rationality requires' is just to comply with all of the *source* requirements of rationality. So knowing what the source requirements of rationality are is knowing *which* rules you are breaking, when you are irrational.[5]

And to this question, Broome grants that there is more than one potential answer. The source requirements of rationality could be *conditional* requirements, requiring

[3] See the appendix to Broome [2007]; also Chapter 10, this volume. The idea is simply this: if rules of rationality tell you to satisfy the material conditional, 'if you believe p, don't believe $\sim p$', then you don't obey all of the rules that apply to you unless you satisfy this conditional. So the wide-scoper accepts this principle. Similarly, according to the narrow-scoper, if you believe p, then there is a rule of rationality that applies to you and tells you to not believe $\sim p$. And so you do not obey all of the rules of rationality that apply to you unless either you don't believe $\sim p$ or this rule does not apply to you—namely, you don't believe p. But that is just to say that you don't obey all of the rules of rationality that apply to you unless you satisfy the material conditional. Even 'myth' theorists (Kolodny [2008a], [2008b]) can agree to this, because they think that everyone is in a situation where either the rules of rationality tell them to believe p or the rules of rationality tell them to not believe $\sim p$. So no one can obey all of the rules of rationality that apply to them without satisfying the same material conditional. None of this should be a surprise, because there is a common piece of data that each of these sorts of views is intended to explain.

[4] Broome [2007].

[5] Broome uses the example of 'the law requires you to' in order to motivate this 'source' sense of the general locution 'X requires you to'. Because Broome doesn't think that 'the law' could be a property, he assumes that we need another way of understanding such expressions, and calls this the 'source' sense of 'X requires you to'. In contrast, I'm not so concerned with whether there is a way of reading 'X requires you to' which lays out what I am calling the 'rules of rationality', or with whether these are appropriately called 'source requriements', as with whether being rational is a matter of obeying them.

specific things of us but only conditionally. That is the narrow-scope view. Or they could be *unconditional* requirements, requiring only that we satisfy some conditional. And that is the wide-scope view. Broome is interested in which of these two views is the more promising view of the source requirements of rationality. That is, in my terminology of rules, he accepts that you count as (having the property of being) irrational in virtue of breaking one or more of the rules (source requirements) of rationality, and he is interested in which rules (source requirements) you are breaking, in interesting cases such as when you have inconsistent beliefs, are akratic or means–end incoherent, or fail to draw the obvious consequences from your other beliefs. So in all of this, we agree about the nature of the question being asked, and about what makes that question interesting.

Broome offers a model for how to think about the logical structure of source requirements. We can think of each source of requirements as yielding a function from world-individual pairs to the set of things that are required by that source of that individual at that world. And Broome assumes that the objects of source requirements—the things that are required—are propositions. So formally, for each source S, each person N, and each possible world w, $R_S(N,w)$ is the set of propositions whose truth is required of N at w. The way that this system allows us to think about conditional requirements, is by allowing the requirement function, R_S, to be non-constant in its world argument. When the members of $R_S(N,w_1)$ are different from the members of $R_S(N,w_2)$, that means that S requires different things of N at w_1 and at w_2. So whenever some proposition *q* belongs to $R_S(N,w)$ whenever the proposition *p* is true at w, we may say that S requires *q* of N, conditional on *p*. There is nothing fancy about this sort of conditional requirement; it is just to say that necessarily, if *p* is true, then S requires *q* of N.

Within Broome's model, the distinction between wide and narrow-scope views about the structure of the rules of rationality that you are breaking when you have inconsistent beliefs is clear: on the wide-scope view, R_S is a constant function whose value includes the proposition that if N believes that *p*, then N does not believe that ~*p*. Whereas on the narrow-scope view, R_S is a non-constant function with the property that $R_S(N,w)$ always includes the proposition that N does not believe that ~*p*, for all values of N, *p*, and w such that N believes that *p* at w. Importantly, both views yield the same predictions about *when* you break the rules of rationality; they simply disagree about which rules you are breaking when you do so.[6]

11.3 Broome's Two Worries

Broome raises two concerns about the narrow-scope view. The first is that "some sources of requirements should not impose inconsistent requirements on you."[7] His

[6] See the appendix to Broome [2007] for further discussion.

[7] Broome [2007, 28].

idea appears to be this: while it may be plausible that you are under conflicting legal requirements, it is not plausible that you are under conflicting requirements of rationality. The rules of rationality are simply not the sort of thing to require inconsistent things of you.

This is a problem, because the narrow-scope view straightforwardly predicts that the rules of rationality which apply to you in a given situation may be inconsistent. Simply suppose that N both believes that *p*, believes that if *p*, then *q*, and also believes that ~*q*, and is clearly considering all three of these beliefs at once. Because N believes that *p* and believes that if *p*, then *q*, a narrow-scope closure rule would require N to believe that *q*. But because N believes that ~*q*, a narrow-scope consistency rule would require N to not believe that *q*. But N cannot both believe that *q* and also not believe that *q*; these requirements are inconsistent. So if rationality requires each of these things of you, then rationality requires inconsistent things. And that, Broome claims, is implausible.

The example that I've just given is only one of a number of ways in which we can combine otherwise plausible narrow-scope rules of rationality in order to generate inconsistent requirements.[8] The key element in all such cases is to start with an agent who is already irrational—plausible narrow-scope views are not going to generate inconsistent requirements of rationality for agents who are fully rational. So the implausible feature of the narrow-scope view is not that because it generates inconsistent requirements, that makes it sometimes impossible to be fully rational. What is supposed to be implausible, Broome claims, is that rationality would require inconsistent things of *anyone*—even someone irrational.

Broome also raises a second worry, though it is less clear. He claims that the source requirements of rationality must obey what he calls the *necessity principle*. According to the necessity principle, if rationality requires you to F, then necessarily, you F if you are rational. In his introduction of the necessity principle, Broome mistakenly suggests that this is a general feature of source requirements. He writes:

> I take it that, if rationality requires you to *F*, then, necessarily, you *F* if you are rational. If morality requires you to *F*, then, necessarily, you *F* if you are moral. And so on. In general, if a source *S* requires of *N* that *p*, then, necessarily, *p* if *N* has the *S*-property. Call this the 'necessity principle'.[9]

But it is easy to see that this is wrong.[10] One of Broome's own leading examples of source requirements is the case of the law. It is well known that many laws are contingent, and could have been otherwise. Indeed, in different jurisdictions, they often are. So if the law requires you to F, it is still perfectly possible for you to be law-abiding and still not F—all that needs to happen, is for the law to be different. Any contingent source of requirements—the law, etiquette, and so on—will therefore lack the necessity property.

[8] Broome [2007] illustrates this point with the norm of non-*akrasia* and the consistency norm on intention, so nothing about the case requires a multi-premise closure rule.

[9] Broome [2007, 33].

[10] Broome now agrees [personal communication] that this was an error.

But Broome doesn't need to assume that the necessity principle applies to all source requirements, in order to have an objection to narrow-scope source requirements of rationality. As with his first argument, all that Broome needs for his argument is that there is something special about *rationality* from which it follows that if there is a source requirement of *rationality*, then necessarily, if you don't comply with it, then you are irrational. In order to support this principle, Broome only needs the view that the source requirements of rationality are themselves necessary.

11.4 The Diagnosis

A single thought can help us to see what makes both of Broome's assumptions seem compelling. In talking of what 'rationality requires of' us, and of what requirements rationality 'imposes on' us, Broome is talking as if rationality—whatever that is—is itself the source of the rules of rationality. And calling the rules of rationality "source requirements" facilitates that way of thinking. This makes it sound like rationality is itself a kind of legislator who creates the rules of rationality and imposes them on us. But in addition to being puzzling exactly what rationality is, so conceived, it is even more puzzling what it does to legislate rules. And if rationality is the legislator, but rationality can't do anything differently, then naturally we should expect its rules to be necessary. Thus we get the idea that the rules of rationality are necessary.

Similarly, if rationality is like a legislator, it is natural to expect it to be a kind of *ideal* legislator. But ideal legislators wouldn't create inconsistent rules—even for people who are already breaking other rules. So the idea that rationality is like an ideal legislator is also just the right kind of idea to motivate Broome's thought that it is implausible to think that rationality ever requires inconsistent things of anyone—even of people who are already irrational.

In contrast, I suggest that a more promising way to think about the narrow-scope picture of the rules of rationality, is as rooted in the idea that *we*—rather than *rationality*—are the authors of the rules of rationality which apply to ourselves. On this picture, it is no wonder that these rules are contingent, because we create them by something that we do. And it is no wonder that they can require inconsistent things of agents who are already irrational, because irrational agents should not be expected to be ideal legislators—on the contrary, it is only to be expected that they would screw up.

Indeed, only on this picture can a narrow-scope account of the rules of rationality serve one of the main purposes on the basis of which I have long advocated narrow-scope principles—namely, to offer an explanation not only of what those rules are, but of where they come from.[11] Part of what is puzzling about wide-scope rules of rationality, with their unconditional application to every rational agent no matter what she is like, is that it is hard to see what explains how these rules get a grip on each

[11] Chapters 7, 8, 9, and 10, this volume.

and every agent. That was one of the two main considerations I offered in Chapter 7, this volume, against wide-scope principles, and the *very same* consideration applies against Broome's picture of narrow-scope rules, on which they are simply conditional rules whose source derives from some external source.

In contrast, if we are the authors of our own rules of rationality, then we can use the conditionality of those laws in order to explain their *source*. This, of course, is exactly what Kant claimed made it easier to explain hypothetical imperatives than to explain the categorical imperative:

> By contrast, 'How is the imperative of morality possible?' is beyond all doubt the one question in need of a solution. For the moral imperative is in no way hypothetical, and consequently the objective necessity, which it affirms, cannot be supported by any presupposition, as was the case with hypothetical imperatives.[12]

The reason that the possibility of hypothetical imperatives does not require any special solution, Kant explains, is that we can appeal to their condition in order to explain them. But since categorical imperatives are unconditional, in their case we can do no such thing. If the condition of narrow-scope rules of rationality is just the condition of their authorship, then we can take advantage of Kant's insight and use that conditionality in order to explain where the rules come from.[13]

It is striking, however, that on this diagnosis, the thought driving both of Broome's arguments against narrow-scope interpretations of the rules of rationality is based on a thought—about who or what is the *source* or *author* of the rules of rationality—that is not captured in his model. In Broome's simple model, the source requirements of rationality simply *are*. The functions in his model make no distinction between worlds in which there are no rules, and worlds in which the rules exist but they are conditional on something false. What I'll now suggest, in what follows, is that if we enrich our model to be able to reflect not just what requirements, laws, or rules there are, but the way in which those requirements, laws, or rules can be *enacted*, and *by whom*, then we will have the resources to distinguish between two very different narrow-scope views, one of which is implausible for all of the reasons that Broome mentions, and one of which offers an attractive picture about not only *what* the rules of rationality are, but of where they *come from*.

11.5 A Model for Legislation

So our goal is to construct a simple way of representing both what the laws are, and how they are enacted. I will take a *legal model* to be a 6-tuple, ⟨*Author, Agent, Enact, Prop*, →, *Legislate*⟩, where *Author* is a set of potential authors of laws, *Agent* is a set of potential agents subject to laws, *Enact* is a set of properties of authors closed under

[12] Kant [2002, 220] (4:419). [13] Compare Chapter 9, this volume.

Boolean operations, *Prop* is a set of properties of agents closed under Boolean operations, and → and *Legislate* are defined as below.

We define *Authority* = {<J,S,D>:J ∈ *Prop*&S ⊆ *Prop*&D ⊆ *Prop*}, and take → to be a relation on *Author* × *Authority*. Intuitively, → tells us which authors have which *authorities*. An *authority* A ∈ *Authority* combines a *jurisdiction* J which tells us the property an agent needs to have in order to fall under that authority with a *scope* S which is the set of things the author is authorized to require of that agent (the set of possible *mandates*), under that authority, and a *discretion* D, which tells us the possible *conditions* of those requirements, under that authority. So intuitively, Y → <J,S,D> means that Y has the authority to require anyone with property J to do anything in S, conditional on any property in D.

→ is a relation, rather than a function, because the same author might have different authorities over different jurisdictions. For example, the president of the United States might have one authority over her children, another authority over the White House staff, and a third authority as commander-in-chief of the United States armed forces. Each of these authorities includes not only a different jurisdiction, but a different scope, and a different discretion. For example, her parental authority may allow her to condition the rules for her children on their gender, but her authority over the White House staff does not. And there are many things her parental authority allows her to require of her children that her authority as commander-in-chief does not allow her to require of the United States armed forces.

Where L = <C,M> ∈ *Law* = *Prop* × *Prop*, we say that L is a (possible) *law*, and call C the *condition* of the law and M its *mandate*. If C ∈ D is consistent with J and M ∈ S, then we say that <C,M> *falls under* authority A = <J,S,D>. Intuitively, this just captures the idea that if you have an authority, then you are authorized to require anything in M of anyone who satisfies J, on any condition in D.

Finally, we take *Legislate* to be a function from *Law* × *Authority* to *Enact*, so that *Legislate* is defined for <<C,M>,<J,S,D>> whenever <C,M> falls under <J,S,D>, and undefined otherwise. Intuitively, the *Legislate* function tells us what the author must do in order to *enact* the law, or in other words, to put it *into force*. So if for some author Y such that Y → <J,S,D>, *Legislate*(<<C,M>,<J,S,D>>)Y, then we say that <C,M> is *enacted* by Y *under* <J,S,D>. If for some Y and some <J,S,D>, <C,M> is enacted by Y under <J,S,D>, we say that <C,M> is *in force*. Obviously, the question of exactly what a given legislator needs to do in order to effect a certain law and why can be an extremely complex question in realistic cases, but for our purposes we may simply treat this as a black box.

All of this sounds like a mouthful, but all that we've done is to introduce the minimal conceptual resources in order to make sense of the idea that laws are created by someone who has the authority to do so, and that there is always something that this legislator must do, in order to create those laws. Just a little bit more terminology will be helpful as we go on. I will say that:

If <C,M> is enacted by Y under <J,S,D>, Jx, Cx, and Mx, then we say that x *complies* with <C,M>. If <C,M> is enacted by Y under <J,S,D>, Jx, Cx, and ~Mx, then we say that x *violates* <C,M>.

If <C,M> is enacted by Y under <J,S,D> and Jx but ~Cx, then we say that x *escapes* <C,M>.
If <C,M> is enacted by Y under <J,S,D> and ~Jx, then we say that x *evades* <J,S,D>.

The distinction between *escaping* a *law* and *evading* an *authority* is not supposed to be a natural-language one between 'escape' and 'evade'; it's a conceptual distinction for which we need a pair of terms, so I choose these.

In order to see how this model works, it will be helpful to walk through a simple example. Consider the following highly simplified case:

Example
Y_1 = New Jersey state legislature $J_1 = \lambda x$(x is in New Jersey)
Y_2 = Oregon state legislature $J_2 = \lambda x$(x is in Oregon)
Y_3 = Wisconsin state legislature $J_3 = \lambda x$(x is in Wisconsin)
S = {P:P is compatible with the United States constitution}
D = {P:P is not a protected category under the "equal protection" clause}
10th Amendment: $\forall n(Y_n \rightarrow <J_n,S,D>)$
C = λx(x is a driver) M = λx(x does not pump her own gas)
Legislate(<<C,M>,$<J_n$,S,D>>) Y_n for n = 1,2 but not 3

The tenth amendment to the United States Constitution authorizes the states to require things of their own residents. Subtleties aside about the exact scope and discretion of this authority, it clearly includes the authority to require drivers not to pump their own gas. And some states have exercised their authority in precisely that way. In particular, the Oregon and New Jersey state legislatures have done whatever it is that they needed to do, in order to enact precisely this conditional law: one that requires you, if you are a driver, not to pump your own gas.

And now we can see our distinctions at work. If you are a driver in New Jersey who does not pump your own gas, then you comply with the law. If you pump your own gas as a driver in New Jersey, then you violate the law. If you are in New Jersey but are not a driver, you escape the law, because it no longer applies to you. And if you leave New Jersey and go somewhere else, then you evade the authority of the New Jersey state legislature, because you no longer fall under the jurisdiction of their authority. But even when you evade this authority, you may not yet escape the law, because if the place you go is Oregon, the same law will apply to you in a different jurisdiction.

11.6 Three Pictures of the Rules of Rationality

Within our model, it is easy to distinguish three distinct pictures of what the rules of rationality say, and where they come from. According to the first, wide-scope, picture, the rules of rationality are universal and unconditional. This is the familiar picture visible in Broome. It looks like this:

Wide
Y = *Rationality* J = λx(x is a rational agent)

S = {P: P supervenes on mental states}
D = S
Y → <J,S,D>
C = *vacuous* M = λx(x does not both believe *p* and believe ~*p*)
Legislate(<C,M>,<J,S,D>) = *vacuous*

On this picture, the source of rational rules is *Rationality*, but rationality does not have to do anything to create those rules, which explains why they are all necessary. Rationality's authority has jurisdiction over all rational agents, and because its requirements are unconditional, every rational agent is bound by every rule of rationality. Among those rules are the requirement not to believe both a proposition and its negation.

A second picture of how the rules of rationality work has much in common with this first picture, but allows that the rules of rationality are narrow-scope. It looks like this:

Narrow, Pass 1
Y = *Rationality* J = λx(x is a rational agent)
S = {P: P supervenes on mental states}
D = S
Y → <J,S,D>
C_1 = λx(x believes *p*) M_1 = λx(x does not believe ~*p*)
C_2 = λx(x believes ~*p*) M_2 = λx(x does not believe *p*)
Legislate(<C_n,M_n>,<J,S,D>) = *vacuous*

On this picture, the author of the rules of rationality is again *Rationality*, but again, Rationality doesn't need to do anything in order to enact its rules, which has the consequence that they are all necessary. However, on this picture, some of them have non-vacuous conditions. In particular, conditional on your having any belief, there is a rule forbidding you to have its contradictory belief. This picture is clearly a narrow-scope picture of what the rules of rationality are like, but because it shares Broome's intuitive thought that the source of the rules of rationality is Rationality itself, it is a consequence of this picture that Rationality may indeed require inconsistent things of us.

However, once we can make distinctions not only about what the laws are and on what condition they apply, but about who enacts the laws and how, we can see that a very different sort of narrow-scope view is intelligible. According to this alternative picture, the rules of rationality are not narrow-scope because rationality imposes conditional requirements on us, but rather because their conditions are the conditions on which we impose requirements on ourselves. This third picture looks like this:

Narrow, Pass 2
$Y_i \in$ *Agent* J_i = λx(x = Y_i)
S = {P: P supervenes on mental states}

$D = S$
$Y_i \rightarrow \langle J_i, S, D \rangle$
C_1 = *vacuous* $M_1 = \lambda x$(x does not believe $\sim p$)
C_2 = *vacuous* $M_2 = \lambda x$(x does not believe p)
Legislate$(\langle C_1, M_1 \rangle, \langle J_i, S, D \rangle) = \lambda x$(x believes p)
Legislate$(\langle C_2, M_2 \rangle, \langle J_i, S \rangle) = \lambda x$(x believes $\sim p$)

On the third picture, there is no need for *Rationality* to be the author of the rules of rationality, because each person is the author of rational rules that apply to herself. Each agent has the authority to require things of herself, and she does so by being in one or another mental state (in this case, at least, by holding a particular belief).

On this third and final picture, the fact that no agent can be rational while believing a contradiction is a fact, not about which rules of rationality there are, but about how an agent legislates a rule of rationality for herself. In believing a contradiction, you both legislate a rule (actually, two rules) of rationality for yourself, and at the same time violate it (them). So the reason that it is impossible to believe a contradiction without breaking the rules of rationality is not that there is any rule of rationality forbidding it, or even that everyone is bound by any rules of rationality at all. It is simply an artifact of what it takes to enact a rule of rationality.

11.7 Commitments

In the abstract, it sounds a bit strange to say that we are the authors of the 'rules' of rationality. And so it is worth exploring just a little bit more what this means. The first thing to note is that on this final picture of the nature of the rules of rationality, the authority under which the rules of rationality are enacted is always of a very special sort: it is an authority whose jurisdiction only includes the author. Such an authority may be helpfully labeled an *autonomous* authority, in the very literal sense that it is a capacity to make laws for oneself. And a helpful name for a law enacted under an autonomous authority is a *commitment*. When, for example, a self-governing legislative body adopts rules of order, it is *committing* to following those rules, even though it has the power to revise them at any point.[14]

Indeed, Sam Shpall [2013], [forthcoming] has independently argued that what I am calling the 'rules' of rationality are more properly called 'commitments'. According to Shpall, commitments are *normative, escapable*, and *agent-dependent*, which fits with the idea that they are autonomous laws—laws whose existence depends on the agent but which may be escaped if the agent does what it takes to no longer legislate that law.

[14] Compare Chapter 10, this volume. Observe how strange it would be to imagine that all self-governing bodies are really regulated by a set of universal conditional rules that are out of their control, instead of seeing them as setting their own rules. My suggestion in this paper is that we make the same mistake about rationality if we imagine it as a set of external rules.

They are also *pro tanto*, which Shpall takes to mean that they allow for conflicts, but *strict* in the sense that failing to abide by them makes one criticizable in a way that is not generally true of failing to act on one's reasons. All of this sounds right about the rules of rationality on our second narrow-scope picture—on which they may conflict (at least for irrational agents) but on which violating any of the rules makes one criticizable as irrational.

With Shpall's terminology in hand, we can re-describe the view in a way that makes it sound much more natural. Rather than saying that the legislative condition for a law that requires one not to believe q is that one believes $\sim q$, we simply say that believing $\sim q$ is a way of committing yourself not to believing q. Similarly, rather than saying that believing p and believing if q then p are what it takes to legislate to oneself a rule requiring belief in q, we simply say that if you believe p and also believe if p then q, then you are committed to believing q. And finally, we can say that when an agent who believes p, believes if p then q, and also believes $\sim q$ is under conflicting rules of rationality, all that this means is that she has inconsistent commitments: she is committed both to believing q and to not believing q.

But all of these claims are independently plausible. All of the time in philosophical conversation we make claims about what one another are committed to believing on the basis of what one another believe, and in full understanding that we have the power to give up those commitments, simply by changing our minds. And we readily recognize—indeed, we often exploit for our argumentative purposes—the fact that people can have conflicting commitments. So if the rules of rationality are autonomous laws, and autonomous laws are just commitments, there is nothing more to the rules of rationality than there is to commitments. And so if irrationality is a matter of violating the rules of rationality, it turns out that avoiding irrationality is simply a matter of living up to your own commitments.[15]

11.8 Conclusion

On the picture that I've been exploring in this paper, there is no mysterious heteronymous source of the rules of rationality; rather, they are rules that we impose on ourselves. This picture answers both of the puzzling features that John Broome found in the idea that the rules of rationality could be narrow-scope, and it exploits the narrow scope of these rules, as I have long advocated doing, to answer a question that is even closer to my heart: how could any heteronymous source of rules gain the kind of authority over every rational agent which we so readily take rationality to have? On the picture that I've described, not only is it much less puzzling how we could have a

[15] Note that Shpall holds that there are both narrow-scope commitments and wide-scope requirements, and so may not accept the thesis that all there is to rationality is living up to your commitments. But my suggestion here is that what you are required to do, in Shpall's sense, is just what you need to do in order to avoid violating one of your commitments.

kind of authority over ourselves, this authority is actually of a very familiar kind: it is simply our ability to *commit* ourselves. And finally, this picture makes clear why it nevertheless seems that the rules of rationality are *not* up to us. That is simply because it is not up to us what it takes to adopt a commitment—only which commitments we adopt.[16]

[16] Special thanks to Errol Lord, Andrew Sepielli, Jonathan Way, Sam Shpall, Alida Liberman, Robert Johnson, Mark Timmons, Baron Reed, and especially to John Broome. Versions of this paper have benefited from audiences at Princeton University Workshop on Normative Theory in October 2010, at Andrew Sepielli's class at the University of Toronto Mississauga in March 2011, and at the University of Southampton in February 2013.

References

Adams, Robert [1973]. 'A Modified Divine Command Theory of Ethical Wrongness'. In Outka and Reeder, eds. [1973], 318–347.

——[1979]. 'Divine Command Metaethics Modified Again'. *Journal of Religious Ethics* 7(1): 66–79.

Bennett, Karen [2004]. 'Global Supervenience and Dependence'. *Philosophy and Phenomenological Research* 68(3): 501–529.

Blackburn, Simon [1973]. 'Moral Realism'. Reprinted in Blackburn [1993], 111–129.

——[1984]. *Spreading the Word*. Oxford: Oxford University Press.

——[1985]. 'Supervenience Revisited'. Reprinted in Blackburn [1993], 130–148.

——[1993]. *Essays in Quasi-Realism*. Oxford: Oxford University Press.

——[1998]. *Ruling Passions*. Oxford: Oxford University Press.

Bond, E. J. [1983]. *Reason and Value*. Cambridge: Cambridge University Press.

Boyd, Richard [1988]. 'How to be a Moral Realist'. In Sayre-McCord, ed. [1988], 181–228.

Brandt, Richard [1979]. *A Theory of the Good and the Right*. Oxford: Oxford University Press.

Bratman, Michael [1981]. 'Intention and Means–End Reasoning'. *Philosophical Review* 90(2): 252–265.

——[1987]. *Intentions, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

——[2009]. 'Intention, Belief, Theoretical, Practical'. In Robertson, ed. [2009], 29–61.

Brink, David [1989]. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Broome, John [1999]. 'Normative Requirements'. *Ratio* 12(4): 398–419.

——[2001]. 'Normative Practical Reasoning'. *Proceedings of the Aristotelian Society*, suppl. vol. 75: 175–193.

——[2004]. 'Reasons'. In *Reason and Value: Themes from the Moral Philosophy of Joseph Raz*. R. Jay Wallace, Philip Pettit, Samuel Scheffler, and Michael Smith, eds. Oxford: Oxford University Press, 28–55.

——[unpublished]. *Reasoning*. Unpublished manuscript, draft of fall, 2005.

——[2007]. 'Requirements'. In *Hommage à Wlodek. Philosophical Papers Dedicated to Wlodek Rabinowicz*. T. Rønnow-Rasmussen, B. Petersson, J. Josefsson, and D. Egonsson, eds. www.fil.lu.se/hommageawlodek

Brown, Campbell [2011]. 'A New and Improved Supervenience Argument for Ethical Descriptivism'. *Oxford Studies in Metaethics* 6: 205–218.

Chisholm, Roderick [1963]. 'Contrary-to-Duty Imperatives and Deontic Logic'. *Analysis* 24(1): 33–36.

Clarke, Samuel [1967] (1706). *A Discourse Concerning the Unchangeable Obligations of Natural Religion, and the Truth and Certainty of the Christian Revelation*. Selections reprinted in D. D. Raphael, ed., *British Moralists 1650–1800*. Oxford: Oxford University Press, 1967, 191–225.

Clarke-Doane, Justin [forthcoming]. 'Moral Epistemology: the Mathematics Analogy'. Forthcoming in *Noûs*.

Cudworth, Ralph [1996] (1731). *A Treatise Concerning Eternal and Immutable Morality*. Sarah Hutton, ed. Cambridge: Cambridge University Press.
Dancy, Jonathan [2000]. *Practical Reality*. Oxford: Oxford University Press.
Darwall, Stephen [1983]. *Impartial Reason*. Ithaca: Cornell University Press.
——[1995]. *The British Moralists and the Internal 'Ought': 1640–1740*. Cambridge: Cambridge University Press.
——[2009]. *The Second-Person Standpoint: Morality and Accountability*. Cambridge, MA: Harvard University Press.
Davidson, Donald [1963]. 'Actions, Reasons, and Causes'. Reprinted in Davidson [1980], 3–20.
——[1978]. 'Intending'. Reprinted in Davidson [1980], 83–102.
——[1980]. *Essays on Actions and Events*. Oxford: Oxford University Press.
Dreier, James [1993]. 'Structures of Normative Theories'. *The Monist* 76(1): 22–40.
——[1996]. 'Rational Preference: Decision Theory as a Theory of Practical Rationality'. *Theory and Decision* 40: 249–276.
Dunaway, William [2013]. 'Realism and Fundamentality in Ethics and Elsewhere'. Ph.D. Dissertation, University of Michigan.
Engstrom, Stephen [1993]. 'Allison on Rational Agency'. *Inquiry* 36(4): 405–418.
Enoch, David [2011]. *Taking Morality Seriously*. Oxford: Oxford University Press.
Ewing, A. C. [1953]. *Ethics*. London: English Universities Press.
Fine, Kit [1994]. 'Essence and Modality'. *Philosophical Perspectives* 8: 1–16.
Foot, Philippa [1975]. 'Morality as a System of Hypothetical Imperatives'. Reprinted in Foot [2002], 157–173.
——[2001]. *Natural Goodness*. Oxford: Oxford University Press.
——[2002]. *Virtues and Vices*. Oxford: Oxford University Press.
Frankfurt, Harry [1971]. 'Freedom of the Will and the Concept of a Person'. *Journal of Philosophy* 68(1): 5–20.
Gensler, Harry [1985]. 'Ethical Consistency Principles'. *Philosophical Quarterly* 35(139): 156–170.
Gibbard, Allan [1990]. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
——[2003]. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
Greenspan, Patricia [1975]. 'Conditional Oughts and Hypothetical Imperatives'. *Journal of Philosophy* 72(10): 259–276.
Hampton, Jean [1998]. *The Authority of Reason*. Cambridge: Cambridge University Press.
Hare, R. M. [1971]. 'Wanting: Some Pitfalls'. In *Agent, Action, and Reason*. Robert Binkley, Richard Bronaugh, and Ausonio Marras, eds. Toronto: University of Toronto Press, 81–127.
Harman, Gilbert [1975]. 'Moral Relativism Defended'. Reprinted in Harman [2000], 3–19.
——[1976]. 'Practical Reasoning'. Reprinted in Harman [1999], 46–74.
——[1978]. 'Relativistic Ethics: Morality as Politics'. Reprinted in Harman [2000], 39–57.
——[1985]. 'Is There a Single True Morality?' Reprinted in Harman [2000], 77–99.
——[1999]. *Reasoning, Meaning, and Mind*. Oxford: Oxford University Press.
——[2000]. *Explaining Value and Other Essays in Moral Philosophy*. Oxford: Oxford University Press.
Hellman, Geoffrey [1985]. 'Determination and Logical Truth'. *Journal of Philosophy* 82: 607–616.
Hermann, Barbara [1993]. *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.
Hill, Thomas [1973]. 'The Hypothetical Imperative'. *Philosophical Review* 82(4): 429–450.

——[1985]. 'Kant's Argument for the Rationality of Moral Conduct'. Reprinted in Hill [1992], 97–122.
——[1989]. 'Kant's Theory of Practical Reason'. Reprinted in Hill [1992], 123–146.
——[1992]. *Dignity and Practical Reason in Kant's Moral Theory*. Ithaca: Cornell University Press.
Hill, Thomas E., Jr. and Arnulf Zweig [2002]. 'Editors' Introduction'. In Kant [2002], 19–108.
Hooker, Brad [1987]. 'Williams' Argument Against External Reasons'. *Analysis* 47(1): 42–44.
Hubin, Donald [1999]. 'What's Special about Humeanism'. *Noûs* 33(1): 30–45.
Huemer, Michael [2005]. *Ethical Intuitionism*. New York: Palgrave Macmillan.
Hughes, G. E. and M. J. Cresswell [1996]. *A New Introduction to Modal Logic*. New York: Routledge.
Jackson, Frank [1998]. *From Metaphysics to Ethics*. Oxford: Oxford University Press.
Joyce, Richard [2002]. *The Myth of Morality*. Cambridge: Cambridge University Press.
Kagan, Shelly [1989]. *The Limits of Morality*. Oxford: Oxford University Press.
Kant, Immanuel [1987] (1790). *Critique of Judgment*. Werner S. Pluhar, trans. Indianapolis: Hackett Publishing Company.
——[1993] (1785). *Grounding for the Metaphysics of Morals (Groundwork)*. James W. Ellington, trans. Indianapolis: Hackett Publishing Company.
——[1996] (1788). *Critique of Practical Reason*. Thomas Kingsmill Abbott, trans. New York: Prometheus Books.
——[1997a] (1785). *Groundwork for the Metaphysics of Morals*. Mary Gregor, trans. Cambridge: Cambridge University Press.
——[1997b]. *Lectures on Ethics*. Peter Heath, trans. Cambridge: Cambridge University Press.
——[2002] (1785). *Groundwork for the Metaphysics of Morals*. Arnulf Zweig, trans. Arnulf Zweig and Thomas E. Hill, Jr., eds. Oxford: Oxford University Press.
Kim, Jaegwon [1984]. 'Concepts of Supervenience'. Reprinted in Kim [1993], 53–78.
——[1987]. '"Strong" and "Global" Supervenience Revisited'. Reprinted in Kim [1993], 79–91.
——[1993]. *Supervenience and Mind*. Cambridge: Cambridge University Press.
Kolodny, Niko [2005]. 'Why Be Rational?' *Mind* 114(3): 509–563.
——[2007]. 'State or Process Requirements?' *Mind* 116(2): 371–385.
——[2008a]. 'Why Be Disposed to be Coherent?' *Ethics* 118(3): 437–463.
——[2008b]. 'The Myth of Practical Consistency'. *European Journal of Philosophy* 16(3): 366–402.
Korsgaard, Christine [1983]. 'Two Distinctions in Goodness'. *Philosophical Review* 92(2): 169–195.
——[1986]. 'Skepticism About Practical Reason'. *Journal of Philosophy* 83(1): 5–25.
——[1996]. *The Sources of Normativity*. Cambridge: Cambridge University Press.
——[1997a]. 'Introduction' to Kant, *Groundwork for the Metaphysics of Morals*. Mary Gregor, trans. Cambridge: Cambridge University Press.
——[1997b]. 'The Normativity of Instrumental Reason'. In *Ethics and Practical Reason*. Garrett Cullity and Berys Gaut, eds. Oxford: Oxford University Press: 215–244.
Lehrer, Keith and Thomas Paxson, Jr. [1969]. 'Knowledge: Undefeated Justified True Belief'. *Journal of Philosophy* 66(8): 225–237.
Lewis, David [1989]. 'Dispositional Theories of Value'. *Proceedings of the Aristotelian Society*. suppl. vol. 63: 113–137.
Mackie, J. L. [1977]. *Ethics: Inventing Right and Wrong*. New York: Penguin.
McGrath, Sarah [2010]. 'Moral Knowledge and Experience'. *Oxford Studies in Metaethics* 6: 107–127.

McLaughlin, Brian [1997]. 'Supervenience, Vagueness, and Determination'. *Philosophical Perspectives* 11: Mind, Causation, and World, 209–230.
McNaughton, David [1989]. *Moral Vision*. Oxford: Basil Blackwell.
Millgram, Elijah [1996]. 'Williams' Argument Against External Reasons'. *Nous* 30(2): 197–220.
——[1997]. *Practical Induction*. Princeton: Princeton University Press.
Moore, G. E. [1903]. *Principia Ethica*. Cambridge: Cambridge University Press.
Nagel, Thomas [1970]. *The Possibility of Altruism*. Princeton: Princeton University Press.
——[1986]. *The View from Nowhere*. Oxford: Oxford University Press.
Oddie, Graham [2005]. *Value, Reality, and Desire*. Oxford: Oxford University Press.
Outka, Gene and John P. Reeder, eds. [1973]. *Religion and Morality: A Collection of Essays*. New York: Anchor Books.
Parfit, Derek [1984]. *Reasons and Persons*. Oxford: Oxford University Press.
——[2011]. *On What Matters*, 2 vols. Oxford: Oxford University Press.
——[forthcoming]. Replies. Forthcoming in *Parfit's Defense of Objectivity in Ethics and Practical Reason*. Peter Singer, ed. Forthcoming from Oxford University Press.
Passmore, J. A. [1951]. *Ralph Cudworth: An Interpretation*. Cambridge: Cambridge University Press.
Paton, H. J. [1947]. *The Categorical Imperative*. Philadelphia: University of Pennsylvania Press.
Paull, Cranston and Ted Sider [1992]. 'In Defense of Global Supervenience'. *Philosophy and Phenomenological Research* 32: 830–845.
Piller, Christian [2007]. 'Ewing's Problem'. *European Journal of Analytic Philosophy* 3(1): 43–65.
Price, Richard [1994] (1748). *A Review of the Principal Questions in Morals*. Facsimile edition. Charlottesville, VA: Ibis Publishing.
Prior, A. N. [1949]. *Logic and the Basis of Ethics*. Oxford: Oxford University Press.
Putnam, Hilary [2004]. *Ethics Without Ontology*. Cambridge, MA: Harvard University Press.
Quinn, Philip [1979]. 'Divine Command Ethics: A Causal Theory'. In *Divine Command Morality*. Janine M. Idziak, ed. New York: Edwin Mellen, 305–325.
——[1990]. 'An Argument for Divine Command Ethics'. In *Christian Theism and the Problems of Philosophy*. Michael Beaty, ed. South Bend: Notre Dame University Press, 289–302.
——[1999]. 'Divine Command Theory'. In *Guide to Ethical Theory*. Hugh LaFollette, ed. Oxford: Basil Blackwell, 53–73.
Raz, Joseph [2005a]. 'The Myth of Instrumental Rationality'. *Journal of Ethics and Social Philosophy*, www.jesp.org 1(1): 2–28.
——[2005b]. 'Instrumental Rationality: A Reprise'. *Journal of Ethics and Social Philosophy*, www.jesp.org, symposium 1: 2–20.
Regan, Donald [1980]. *Utilitarianism and Cooperation*. Oxford: Oxford University Press.
Robertson, Simon, ed. [2009]. *Spheres of Reason: New Essays in the Philosophy of Normativity*. Oxford: Oxford University Press.
Ross, Alf [1941]. 'Imperatives and Logic'. *Theoria* 7(1): 53–71.
Ross, Jacob [2006]. 'Acceptance and Practical Reason'. Ph.D. Dissertation, Rutgers University.
Sayre-McCord, Geoffrey [1988]. *Essays on Moral Realism*. Ithaca: Cornell University Press.
Scanlon, T. M. [1998]. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
——[2009]. *Being Realistic About Reasons*. Locke Lectures. Last accessed from Oxford University philosophy faculty website at http://www.philosophy.ox.ac.uk/__data/assets/pdf_file/0008/25856/Lecture_2.pdf on 1/23/2013.
Scheffler, Samuel [1982]. *The Rejection of Consequentialism*. Oxford: Oxford University Press.

Schiffer, Stephen [1987]. *Remnants of Meaning*. Cambridge, MA: MIT Press.
Schmitt, Johannes and Mark Schroeder [2011]. 'Supervenience Arguments Under Relaxed Assumptions'. *Philosophical Studies* 155(1): 133–160.
Schneewind, J. B. [1998]. *The Invention of Autonomy*. Cambridge: Cambridge University Press.
Schroeder, Mark [2005a]. 'Instrumental Mythology'. *Journal of Ethics and Social Philosophy*, www.jesp.org, symposium 1, 2–12.
——[2005b]. 'Realism and Reduction: The Quest for Robustness'. *Philosophers' Imprint* 5(1): www.philosophersimprint/005001
——[2007a]. *Slaves of the Passions*. Oxford: Oxford University Press.
——[2007b]. 'Teleology, Agent-Relative Value, and "Good"'. *Ethics* 117(2): 265–295.
——[2007c]. 'Weighting for a Plausible Humean Theory of Reasons'. *Noûs* 41(1): 138–160.
——[2008]. 'Having Reasons'. *Philosophical Studies* 139(1): 57–71.
——[2009a]. 'Buck-Passers' Negative Thesis'. *Philosophical Explorations* 12(3): 341–347.
——[2009b]. 'A Matter of Principle'. Joint critical notice of Jonathan Dancy, *Ethics Without Principles*, and Sean McKeever and Michael Ridge, *Principled Ethics*. *Noûs* 43(3): 568–580.
——[2010a]. 'How to be an Expressivist About Truth'. In *New Waves in Truth*. Nikolaj Jang Pedersen and Cory Wright, eds. New York: Palgrave Macmillan, 282–298.
——[2010b]. *Noncognitivism in Ethics*. New York: Routledge.
——[2011]. '*Ought*, Agents, and Actions'. *Philosophical Review* 120(1): 1–41.
Setiya, Kieran [2004]. 'Hume on Practical Reason'. *Philosophical Perspectives* 18 (Ethics): 365–389.
——[2007a]. 'Cognitivism About Instrumental Reason'. *Ethics* 117(4): 649–673.
——[2007b]. *Reasons Without Rationalism*. Princeton: Princeton University Press.
Shpall, Sam [2013]. 'Wide *and* Narrow Scope'. *Philosophical Studies* 163(3): 717–736.
——[forthcoming]. 'Moral and Rational Commitment'. Forthcoming in *Philosophy and Phenomenological Research*.
Sidgwick, Henry [1981] (1907). *The Methods of Ethics*. Indianapolis: Hackett.
Smith, Michael [1994]. *The Moral Problem*. Oxford: Basil Blackwell.
——[2003]. 'Neutral and Relative Value After Moore'. *Ethics* 113(3): 576–598.
Streumer, Bart [2008]. 'Are There Irreducibly Normative Properties?' *Australasian Journal of Philosophy* 86(4): 537–561.
——[2011]. 'Are Normative Properties Descriptive Properties?' *Philosophical Studies* 154(3): 325–348.
Tenenbaum, Sergio [2007]. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge: Cambridge University Press.
Toulmin, Stephen [1950]. *Reason in Ethics*. Cambridge: Cambridge University Press.
van Roojen, Mark [unpublished]. 'Consequents of True Practical Conditionals Detach'. Unpublished paper.
Wallace, Jay [2001]. 'Normativity, Commitment, and Instrumental Reason'. *Philosophers' Imprint* 1(3): http://www.philosophersimprint.org/001003
Watson, Gary [1975]. 'Free Agency'. *Journal of Philosophy* 72(8): 205–220.
Way, Jonathan [2010]. 'Defending the Wide-Scope Approach to Instrumental Reason'. *Philosophical Studies* 147(2): 213–233.
——[2012]. 'Explaining the Instrumental Principle'. *Australasian Journal of Philosophy* 90(3): 487–506.
Wedgwood, Ralph [1999]. 'The Price of Non-reductive Moral Realism'. *Ethical Theory and Moral Practice* 2(3): 199–215.

——[2000]. 'The Price of Non-Reductive Physicalism'. *Noûs* 34(3): 400–421.
——[2007]. *The Nature of Normativity*. Oxford: Oxford University Press.
Williams, Bernard [1981]. 'Internal and External Reasons'. In *Moral Luck*. Cambridge: Cambridge University Press, 101–113.
Wilson, Jessica [2010]. 'What is Hume's Dictum, and Why Believe It?' *Philosophy and Phenomenological Research* 80(3): 595–637.

Index

actualism *see* possibilism
Adams, Robert 34n16
agent-neutrality 150–2, 157–8, 160, 163, 167, *see also* reasons, agent-neutral
akrasia 187, 188–9
analyticity 206–10
authority 20, 23–5, 233–7
autonomy 13–14, 212–13, 223, 224–6, 237–9

Balguy, Richard 3
Bennett, Karen 125n4
Blackburn, Simon 112–13, 125n3, 202n5
Bond, E.J. 60n1, 64n10, 151
Boyd, Richard 38n25
Brandt, Richard 212
Bratman, Michael 174n3, 191n27
Brink, David 38n25
Broad, C.D. 3
Broome, John 10, 13, 48n7, 148n2, 148n3, 150n8, 153–4, 171, 174n2, 176, 190n26, 193n31, 196, 202n5, 211n15, 219, 227–33, 235–7
Brown, Campbell 125n5
Bykvist, Krister 137n17

Calvinism 19
Clarke, Samuel 3, 15, 20n2, 21n4, 29n9, 38, 138n19
Clarke-Doane, Justin 124n1
Commitments 237–9
conscience 159–61, *see also* Ewing's Problem
consequentialism 35–7, 48–9, 68n12, 95
conservativism 77–8, 86–9, 92–5
Cresswell, M.J. 97
Cudworth, Ralph 3–4, 15, 19–41
Cudworthy argument 3–4, 21–6, 28, 37–39, 138

Dancy, Jonathan 148n2, 150n7
Darwall, Stephen 24n7, 36n19, 37n20, 60n1, 64n10, 71n16, 148n2, 148n3, 170–2, 184n24, 201, 202n4, 211n15, 212, 219
Davidson, Donald 55n15, 192n29
Descartes 19
Dreier, James 35n17, 35n18, 202n5
Dunaway, William 110n11, 126n6

Engstrom 202n3
Enoch, David 124n1
Epicurus 19
Ewing's Problem 174–80
explanations 1–2
  constitutive 3, 7, 33–4
  explanatory arguments *see* supervenience, explanatory arguments
  non-normative 26–7, 32–4
  normative 19, 25–6, 27–30, 139–40

Foot, Philippa 4, 62–3, 64
Frankfurt, Harry 212

Gensler, Harry 148n2, 174n2, 176n7, 202n5, 211n15
Gibbard, Allan 212
goodness, instrumental, *see* instrumental goodness
goodness, intrinsic, *see* intrinsic goodness
Greenspan, Patricia 148n2, 171n26, 177, 202n5

Hampton, Jean 5, 14, 19, 38–9, 31n13, 38, 148n2, 152, 202n4, 202n5, 204, 217n4
Hare, R.M. 148n2
Harman, Gilbert 2, 4, 62–3, 64, 196
Hellman, Geoffrey 101n6
Hermann, Barbara 202n4
Hill, Thomas 11, 148n2, 174n2, 201, 202n4, 205, 207, 209, 211n15, 214, 216–18, 219, 225n18, 226
Hobbes, Thomas 19
Hooker, Brad 60n1, 64n10
Hubin, Donald 60n1, 64n9
Huemer, Michael 90n7, 128n9
Hughes, G.E. 97
Hume, David 61n4
Hume's Dictum 126–8, 132, 137, 138, 142–3
Humean Theory of Reasons 60–79, 152, 170–2
  broad 61, 70
  classical argument for 63–4
  narrow 61
hypothetical imperatives *see* Kant, Immanuel on hypothetical imperatives

instrumental goodness 2
instrumental rationality, *see* rationality, instrumental
intrinsic goodness 2, 7, 140

Jackson, Frank 96, 103, 105–6, 111–12, 125–6, 133n13
Joyce, Richard 4, 38n24, 63n8
jurisdiction 12–13, 218, 222–4

Kagan, Shelly 35n17, 35n18, 39n9
Kant, Immanuel 5, 15, 46–7, 199
  on categorical imperatives 203–5, 212–13
  on hypothetical imperatives 5, 11–12, 145, 160n20, 201–15, 216–18, 224–6, 233
Kim, Jaegwon 96, 101n6, 102–3, 103–5, 111–12, 116n14, 125–6, 133n13
Kolodny, Niko 176n8, 219, 229n3
Korsgaard, Christine 11, 14, 19, 31n12, 38–9, 45n4, 60n1, 61n4, 64n10, 69, 202n4, 208–10, 211n15, 214
laws 12–13, 217–18, 221–4, 233–5
Lewis, David 212

McGrath, Sarah 124n1
Mackie, J.L. 4, 63, 64, 84
McLaughlin, Brian 117n16
McNaughton, David 40n26
McPherson, Tristram 120n19
mathematics 128–30
means-end coherence 173–4, 180–2, 184–7, 189–92, 196–7
mere permissibility *see* picking
Mill, John Stuart 15
Millgram, Elijah 64
miners problem *see* three envelope problem
modal logic 9, 96–7
Moore, G.E. 3, 6–7, 15, 36, 38n25, 140, 141n24, 142n25
  paradox, *see* Moorean contradiction
Moorean contradiction 47
Mrongovious 205

Nagel, Thomas 40n26, 42n2, 53n11, 55n15, 68n12, 128n9
narrow scope *see* wide scope
naturalism 86n4, 89n6
Nietzsche, Friedrich 84
non-cognitivism 83n2, 93

Ockham 19
Oddie, Graham 40n26
open question argument, 7, 36, 39–40
ought 50n10, 147–8, 149, 152–4, 201–2, 219–20
  epistemic 174
  subjective 178–80, 192–4

Parfit, Derek 7, 40n26, 81, 83–95, 124n1, 128n9, 193n30
Passmore, J.A. 24n7
Paton, H.J. 202n4
Paull, Cranston 101n6, 117n16
picking 187–8
Piller, Christian 176n6
Plantinga, Alvin 94
pluralism 7
possibilism 94
Price, Richard 3, 9, 15, 17, 20n2, 21, 29n10, 39, 124–5, 129, 138–41, 143–4
Prichard, H.A. 3
principles 214–15
Prior, Arthur 24n7, 30n10
Protagoras 19
Putnam, Hilary 124n1

Quinn, Philip 34n16

rationalism 45–6, 208n11, 208n12
rationality 14
  instrumental, 4–5, 147–59, *see also* means-end coherence
Rawls, John 11, 202n4
Raz, Joseph 174n3, 184n24, 191n28, 196n33, 219
realism
  non-reductive 124, 131–2
reasons 5–6, 154–5, 194–7
  agent-neutral 6, 43–59
  agent-relational 6, 43–5, 65, 70–1
  Humean theory of *see* Humean Theory of Reasons
  subjective, 158–9
  weight of 92, 195
reductivism 2, 7–9, 34, 37–41, 78–9, 86–92, 96, 102–6, 108–12
Regan, Donald 193n30
requirements 227–30, *see also* laws
revisionism 77
Ross, Jacob 193
Ross, W.D. 3, 15, 83
Rules 227–8, 232–3, 235–7, 238–9

S5 9, 96–7
Scanlon, T.M. 8, 42n1, 125n3, 134–7, 142, 143–4, 151
Scheffler, Samuel 49n9
Schiffer, Steven 125n3
Schmitt, Johannes 8–9, 96–123, 128, 143n26
Schneewind, J.B. 24n7, 36n19
Schroeder, Carol 126
Setiya, Kieran 61n4, 174, 177n9, 181n16, 182n17, 196
Shpall, Sam 224, 237–8
side effects 190–1
Sider, Ted 101n6, 117n16
Sidgwick, Henry 7, 15, 83, 129, 148n2
skepticism, moral 62–3
Slaves of the Passions 6, 14
Smith, Michael 5, 49n9, 212
Standard Model 2–5, 27–30, 71–4, 138–41
Standard Model Theory 6, 28, 30–1, 74, 138–41, 143
Streumer, Bart 125n5
Sturgeon, Nicholas 2
subsumption 52–9, *see also* Standard Model Theory

supervenience 8–9, 96, 98–102, 102–15, 115–21, 124
and covariation 129–30
direct arguments for reduction 102–6, 125–6, 134
explanatory arguments for reduction 108–10, 125–8
global 96, 100–02, 105–6, 114–15
specific supervenience facts 107–8
strong 96, 99–100, 102, 103–5, 115–18, 121–3, 133
symmetry 149–50, 156–7, 159, 162–3, 165–7, 176–7

Tenenbaum, Sergio 192n29
three envelope problem 187, 192–5
Toulmin, Stephen 193n31
transmission 183–4, 195–6
triviality objection 89–92

vegetarianism 83
voluntarism 19–21, 34, 38–9

Wallace, Jay 148n2, 174n2, 188, 196
Watson, Gary 212
Way, Jonathan 11, 219
Wedgwood, Ralph 8–9, 96–7, 112–15, 128, 143n26
wide scope 10, 147–8, 149–52, 156–7, 159–60, 161–4, 164–7, 176–8, 219–20, 227–9, 230–2, 235–7
and Kant 201–5, 206–10, 211–12, 214
Williams, Bernard 60n1, 64n10, 84, 118–21, 178n12, 212
Williams, Nate 161n21
Wilson, Jessica 126n7

Zweig, Arnulf 217n3