

©2018
Pamela Grace Robinson
ALL RIGHTS RESERVED

Toward an Ultimate Normative Theory

By

Pamela Grace Robinson

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Andy Egan

And approved by

New Brunswick, New Jersey

October 2018

ABSTRACT OF THE DISSERTATION

Toward an Ultimate Normative Theory

By PAMELA GRACE ROBINSON

Dissertation Director:
Andy Egan

In this dissertation I take up two challenges for any final or ultimate normative theory: that of explaining the apparent conflict between different normative domains, and that of accommodating all subjective normative facts. I propose a framework for a normative theory that explains how conflict between normative domains could be consistent, and I describe a way to build all-things-considered normative concepts that take all domains into consideration. However, I argue that the second challenge is more serious: that no normative theory could accommodate all subjective normative facts. If I'm right, where does this leave us? I propose a way to extend the solution to the first challenge to accommodate a very large number of subjective normative facts, and I suggest that we think of this theory as an 'open-ended' one that's meant to be extended indefinitely.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of many people. I am especially grateful to my advisor, Andy Egan, for his consistent support and for his many careful comments and keen insights.

For fruitful discussions and/or comments on drafts, I thank: Amanda Askill, Sam Carter, Charles Côté-Bouchard, Ruth Chang, Marilie Coetsee, Miguel Egler, Adam Elga, Megan Feeney, Claire Field, Stephen Finlay, Brandon Fitelson, Danny Foreman, Adam Gibbons, Jimmy Goodrich, Simon Goldstein, Verónica Gómez, Alan Hájek, Josef Holden, Tyler John, David Anton Johnson, Matthew Kopec, Stephanie Leary, Patrick Miller, Andrew Moon, John Phillips, Daniel Rubio, Ernest Sosa, Larry Temkin, Peter van Elswyk, Christopher Willard-Kyle, and Daniel Wodak. I also thank audiences at the Humboldt Foundations of Normativity Workshop, the Arché Graduate Conference, and the Saint Louis Annual Conference on Reasons and Rationality.

I am indebted to Holly Smith and Alvin Goldman for their encouragement with this project throughout graduate school. David Black and Philip Swenson also deserve special mention for spending so much time discussing subjective normativity with me. Chapters 2 and 3 are in a large part my responses to these discussions.

I thank my family for always accepting and understanding my life choices. Most importantly, I thank Justin Kalef for now, as ever, inspiring me to be a better thinker and a better person.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
List of Tables	vi
List of Illustrations	vii
Introduction	1
1. Normativity <i>Lite</i>	5
1.1. Apparent Conflict	7
1.2. Normativity <i>Lite</i>	14
1.2.1. The Normativity <i>Lite</i> Framework for SHOULD	15
1.2.2. The General Normativity <i>Lite</i> Framework	20
1.3. An All-Things-Considered SHOULD	22
1.4. An Adequate Version of Pluralism	25
1.5. Monism <i>versus</i> Pluralism	30
2. The Incompleteness Problem for Normativity	34
2.1. The Incompleteness Argument	34
2.2. Objections and Ways Out	46
2.2.1. ‘Optimistic’ Responses	47
2.2.1.1. Hierarchy Theories	47
2.2.1.2. Other Kinds of Complexity	51

2.2.2. ‘Fastidious’ Responses	57
2.2.2.1. Answers	57
2.2.2.2. Questions	60
2.2.3. ‘Dismissive’ Responses	61
3. Starting an Open-Ended Normative Theory	64
3.1. The Idea of an ‘Open-Ended’ Normative Theory	65
3.2. A Good First Stage in an Open-Ended Normative Theory	71
3.3. Infinite Hierarchy Theories as Good First Stages	75
3.4. Proposal for a Good First Stage: the <i>Lite</i> Hierarchy Theory	83
3.5. Two Applications	97
3.5.1. The Problem of Moral Uncertainty	98
3.5.2. The Problem of Higher-Order Defeat	100
3.6. Next Stages	107
Conclusion	113
Appendices	115
A. Adjusting the Normativity <i>Lite</i> Schema	115
B. Relationalism and Finlay’s End-Relationalism	121
C. Uncertainty	127
D. The Incompleteness Problem and Action-Guidance	129
E. Extending the <i>Lite</i> Hierarchy Theory in Other Directions	137
Bibliography	140

LIST OF TABLES

Table 1. How the strategies for responding to cases of apparent conflict meet the desiderata	13
Table 2. Decision table for Parfit's Miners Case	37
Table 3. Decision table for MacAskill's Chimpanzee Case	41
Table 4. Reachable states of affairs	85
Table 5. Reachable states of affairs ranked according to value	85
Table 6. How Jill might complete a table showing her reachable states of affairs	87
Table 7. How Jill 2 might complete a table showing her reachable states of affairs	89
Table 8. How Jill 2 might (more carefully) complete a table showing her reachable states of affairs	89
Table 9. How Jill might (more carefully) complete a table showing her reachable states of affairs	90
Table 10. How Jill 3 might complete a table showing her reachable states of affairs	93

LIST OF ILLUSTRATIONS

Figure 1. Three strategies for responding to cases of apparent conflict	9
Figure 2. Normative space covered by an incomplete theory	65
Figure 3. Normative space covered by an ultimate theory	66
Figure 4. Normative theories that cover successively more space	67
Figure 5. An open-ended normative theory	69
Figure 6. An open-ended normative theory with two stages	69
Figure 7. A closed two-level normative theory	70
Figure 8. An open-ended two-level normative theory	71
Figure 9. A good first stage in an open-ended normative theory	72
Figure 10. A worse first stage in an open-ended normative theory	72
Figure 11. An even worse first stage in an open-ended normative theory	72

INTRODUCTION

The Idea of an Ultimate Normative Theory

Normative theorists engage in many different projects. Moral theorists construct theories that tell us what our moral obligations are, what we morally ought to do, what is morally good, and so on. Epistemologists build theories that tell us when our beliefs or credences are justified or epistemically rational, what is epistemically valuable, and so on. Decision theorists work on rules for rational decision-making. Legal theorists, grammarians, writers on etiquette, game designers, and many others also work with, theorize about, and create, normative facts. Normative facts concern how things should be or ought to be as opposed to how they are. Or they are evaluative: *that was the best pie, that is a good knife, this will be one of the worst winters in decades*. Some normative theorists theorize about what kinds of facts really belong in the category of the ‘normative’ as opposed to something else. But however the boundaries are drawn, it’s reasonable to ask whether there could be a final, all-encompassing, or *ultimate* normative theory that ‘covers’ all normative facts.

Imagine that normative facts comprise a two-dimensional normative *space*. Normative theorists work to cover or *map* various portions of that space. A reasonable question, then, is whether we could find a single map or even a collection of maps that accurately cover the whole of normative space. Finding such a theory could be seen as the final aim of normative theorists.

This dissertation starts with the assumption that finding or building an ultimate normative theory is the end goal of all our theorizing. It confronts two of the biggest challenges to such a project, and suggests ways of making progress. The first is the problem

of unifying different normative ‘domains.’ For example, the moral, the epistemic, the prudential, the legal, and so on. Can these fit in a single map? Or, do we need a separate map for each, and how could this be done? The second problem is finding a theory that could cover all the ‘subjective’ normative facts. These facts make reference to our own imperfect epistemic states—for example, to our uncertainty about things. And since there is almost no limit to what we can be uncertain about, there threaten to be too many subjective normative facts to be covered by any single map or even by any collection of maps.

Chapter 1: Normativity Lite

In Chapter 1, I take up the first challenge and propose the Normativity *Lite* Framework to explain how it can be resolved. The Normativity *Lite* Framework doesn’t tell us whether all the different normative domains could be unified or not. But it does offer a story of how, if this were possible, they could be joined with a single ‘all-things-considered’ theory that tells us what we should do when different kinds of normative considerations come into conflict. It also explains why we might need a number of independent maps of normative space, and how this could work.

The Normativity *Lite* Framework divides normative space into two portions: facts that involve evaluations of the options that normative agents have, and facts that don’t. Then it populates the first portion of normative space with minimal or ‘*lite*’ normative facts, which are extremely numerous and easy to generate. All normative facts in this first portion are given an analysis that makes them relative to states of affairs. Facts about how valuable states of affairs are determine which *lite* normative facts count as full-fledged, important, or even ‘all-things-considered’ ones.

In this way, the Normativity *Lite* Framework can generate different maps of normative space (even collections of maps as opposed to single ones), depending on the theory of value that's fed into it. Our theory of value ends up explaining the extent to which our ultimate normative theory can be unified.

Chapter 2: The Incompleteness Problem for Normativity

In Chapter 2, I pose a problem for the project of building an ultimate normative theory. The trouble, I argue, is that no normative theory could be complete. The argument goes as follows. It's possible to be uncertain about any normative theory. And if it's possible to be uncertain about a normative theory, then we can ask the question: *what should one do given one's uncertainty about it?* But no normative theory could answer this question, when the uncertainty is about itself. And a complete normative theory would be able to answer every normative question. So no normative theory could be complete.

I consider three kinds of objections to this argument: 'optimistic' ones that attempt to find a normative theory that *could* answer the question *what should one do given one's uncertainty about it?*, 'fastidious' ones that attempt to find flaws in the original formulation of the argument, and 'dismissive' ones that attempt to downplay the conclusion. None of them, I argue, are very successful.

Continuing with the mapping metaphor, the incompleteness problem arises because subjective normative facts about what agents should do can't be mapped by just any normative theory. They must be mapped by theories that are sensitive to the epistemic shortcomings of the agents. And in the case of uncertainty, this means theories that don't rely on the truth about what the agent is uncertain about. Because no theory can be neutral

about itself in this way, no theory could map all of normative space or answer every normative question.

Chapter 3: Starting an Open-Ended Normative Theory

In Chapter 3, I attempt to address the challenge posed by Chapter 2. If it's impossible to ever find or build an ultimate normative theory, then what should our aim be instead? If any normative theory would leave some part of normative space uncovered, what kind of project should we engage in?

I suggest that we build an 'open-ended' normative theory: one that can be extended indefinitely. Such a theory would be built in stages. The first stage would answer many normative questions, or cover a good amount of normative space, but would leave some subjective questions unanswered. The second stage would answer these questions, but leave some even more subjective questions to be answered, and the third stage would answer those, and so on.

I also propose a first stage in such a theory: the *Lite* Hierarchy Theory. This theory appeals to the Normativity *Lite* Framework in Chapter 1, and describes a way to extend it to produce an infinite number of more and more subjective kinds of normative concepts. The theory also promises to answer more normative questions than its rivals. Like the Normativity *Lite* Framework, it's more of a blueprint of a theory than a theory, so the project of filling in the details could be a worthy and 'provisionally ultimate' goal for normative theorists. As it leaves some normative space uncovered, it also offers a glimpse of some of the challenging tasks that lie ahead.

1

Normativity *Lite*

Suppose you're deciding whether to buy a new car or donate to charity. What should you do? We might say: *prudentially* you should buy the car, and *morally* you should give to charity. But putting things this way raises questions. For example, how are you to decide what to do if knowing these facts about 'what you should do' doesn't settle it?

This is an example of a kind of *apparent conflict* that troubles normative theorists. Cases of 'apparent conflict,' as I'll call them, are those in which it seems that (in some sense) one thing should be done, and also that (in some other sense) some other thing should be done instead. The cases I consider here involve a disagreement between different 'normative domains.'^{1,2} For example, between what's morally best, prudentially best, epistemically best, and so on.

In this chapter I survey different strategies for resolving apparent conflict and offer a theory according to which either of the two best ones could work. These most promising strategies are *sophisticated monism*, according to which there's a single 'all-things-considered' fact about what should be done, and *pluralism*, according to which there are many different

¹ Or cases in which it seems that (in some sense) one thing is *best* or *rational* or..., but also that (in some other sense) some other thing is *best* or *rational* or... instead. I'm also not using 'sense' as a technical term here, but just to convey the intuition.

² Another kind of case arises when there's divergence between something like 'what's objectively best' and 'what's subjectively best.' I discuss cases like this in Chapter 2, though not with the aim of explaining how this kind of apparent conflict could even make sense. Unlike the first, this second kind of apparent conflict is no impediment to action. If you're ever in the position know exactly what you objectively should do and also what you subjectively should do, then what you subjectively should do will be the same as what you objectively should do.

and consistent facts about what should be done.³ The theory I propose is ‘the Normativity *Lite* Framework.’ It posits a vast number of minimally normative concepts, as well as a way to generate any (fully) normative concept we might need. I explain how it can be used to offer the all-things-considered concepts required by the sophisticated monist as well as the different kinds of should concepts required by the pluralist. I also use the framework to locate the main point of disagreement between monists and pluralists, which turns out not to be about the number of normative concepts but about the possibility of a privileged all-things-considered one.

I’ll conclude that the Normativity *Lite* Framework can help us make significant progress toward resolving cases of apparent conflict. If we use it, we can be sure that there will be an acceptable solution, and we’ll know what the possibilities are and what would decide between them.

A note about methodology: this is a *potentially revisionary* conceptual project.⁴ My aim is the normative one of *finding the best way to organize normative conceptual space* as opposed to the descriptive one of *discovering the nature of the normative concepts we already have*.⁵ While there’s little agreement about exactly how to carry out each kind of project, most can agree that what’s evidence for a descriptive theory may not be evidence for a potentially revisionary one, and *vice versa*. For example, if this project were a descriptive one, I’d claim that the theory I propose correctly identifies and analyzes the normative concepts we already have. And to back that up, it may help to appeal to linguistic analysis or data about folk intuitions.

However, that’s not my aim, and the method I’ll use is different. It’s roughly this: *determine*

³ I have chosen to define these views in a simple way that’s in line with how people usually think about them as opposed to a way that makes them inconsistent. I also define more precise versions of each in section 1.4.

⁴ If there are no such thing as concepts, then what I say here could be translated into talk of something else; for example: relations, properties, or expressions in an ideal language.

⁵ The project is ‘potentially’ revisionary because I don’t assume that our current system is suboptimal.

what we theorists want or expect from an ideal normative theory, as well as the constraints we think apply to it, and then provide a theory that can deliver as many of these things as possible while meeting the constraints.

1.1. Apparent Conflict

I'm about to present some cases of apparent conflict. You may not find these cases deeply puzzling. You can probably answer the questions about what the agents should do and make your answers sound reasonable. Nevertheless, cases like these raise difficult questions about the *structure* of normative conceptual space. What normative theorists find puzzling is not so much the cases themselves as *what normative conceptual space would have to be like in order for the questions to make sense.*

Abe is diagnosed with an illness and learns that his odds of recovery are 1 in 10. He then learns that he'll triple his odds of recovery if he *believes* that he'll recover. Should Abe believe that he'll recover, or not?⁶

Bess is deciding whether to order a beef burger or a veggie burger. She prefers beef, but believes that vegetarianism is morally required. Should Bess order the beef or the veggie burger?⁷

⁶ Some normative theorists think that we don't have direct control over what we believe. (See, for example, Ginet 2001.) If you hold a view like this, then it may seem that the case should be re-described so that Abe must choose whether or not to take a belief pill that would produce the belief that he will recover. And, it might be thought, in a case like this there's no conflict between prudential and epistemic considerations: prudential considerations are the only ones that matter. However, it's not obvious that there's no apparent conflict even in this case. It's still plausible that in *some* epistemic sense, it would be wrong of Abe to take the pill. (Also see Rinard 2017 for an argument that it may not matter much whether we only have indirect control over our beliefs or not.)

Carl has been forced to play chess with a vicious and insecure mob boss, Griselda. Carl can see that he'll win if he plays *rook to E8* and that he'll lose if he plays *bishop to E8*. But he's also certain that Griselda will be in a murderous mood if he wins, whereas she'll be in a generous one if he loses. Should Carl play his rook or his bishop?⁸

In each case, two different kinds of considerations seem to bear on what should be done.⁹ And they seem to support opposing answers. However, Abe, Bess, and Carl can each only do one thing. Pre-theoretically, what these cases suggest is that an ideal normative theory would:

Two Desiderata for an Ideal Normative Theory

1. Accommodate the thought that there's *some* kind of relativity here, if only because there are two kinds of considerations that genuinely support different conclusions.
2. Accommodate the thought that our theory must privilege one option as something like 'the-thing-to-be-done,' since each agent can only do one thing.

And there are three main strategies for responding to such cases:¹⁰

⁷ This case might be read as an example of *akrasia*: Bess believes that she shouldn't order the beef burger but wants to do it anyway. However, this is not the kind of conflict the case is meant to illustrate. We can imagine that Bess has no weakness of will and is perfectly able to order the vegetarian option. She can see that one choice is the moral one and that the other would best satisfy her craving, and genuinely wonders which one she should choose.

⁸ This case might be read as one of conflicting prudential considerations: Carl wants to win but also wants to stay alive. However, it's easier to see the kind of conflict intended if we imagine that Carl has no desire whatsoever to win the game. Even if all he cares about is his own safety, it still seems that there's something good to be said in favor of playing *rook to E8*—after all, it's the *winning move*.

⁹ You can think of 'considerations' as 'reasons.' I'm avoiding the use of the term 'reason' here because I'm imagining that we're approaching these cases pre-theoretically.

¹⁰ To simplify my inquiry, I don't consider 'mixed' views according to which some domains are treated differently than others. For example, one might want to keep the practical and theoretical domains separate while positing a single all-things-considered kind of normativity for each domain.

The *simple monist* strategy: The answers are inconsistent and one is true. One kind of consideration at most *apparently* counts in favor of the other answer.¹¹

The *sophisticated monist* or *all-things-considered* strategy: The answers are inconsistent and one is true. Different kinds of considerations genuinely count in favor of each answer, but one answer is favored over all.

The *pluralist* strategy: The answers are consistent and both are true.

In a picture:

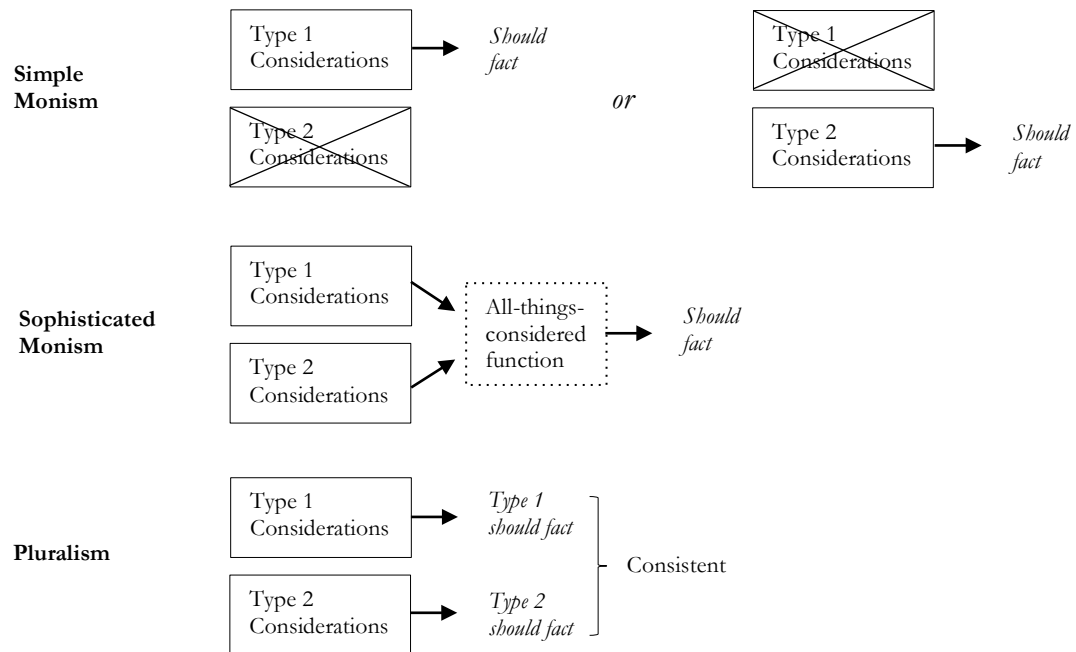


Figure 1. Three strategies for responding to cases of apparent conflict¹²

Such a view would count here as a version of pluralism, but it would import large parts of sophisticated monism.

¹¹ Note that it's possible to hold this view while recognizing the existence of different normative domains, so long as they could never come into conflict.

¹² In this diagram, the 'should facts' stand for answers. The six possible (unqualified) should facts are: *Abe should believe that he'll recover*, *Abe should not believe that he'll recover*, *Bess should order the beef burger*, *Bess should order the veggie burger*, *Carl should move his rook*, *Carl should move his bishop*. The pluralist posits 'types' of should facts, so she'll recognize a larger set of should facts.

In this section I'll compare how promising each strategy looks at the outset, imagining that we've been presented with cases of apparent conflict for the first time and that we don't yet have complex theoretical tools on hand.

Monists give one answer to each case. For example, in Abe's the choice is between:

A1. Abe should believe that he'll recover.

A2. It's not the case that Abe should believe that he'll recover.

What distinguishes the *simple* from the *sophisticated* monist is how she supports her answers. Some simple monists will give answer A1, noting that it's in Abe's best interest to form the belief. Those following Clifford, who held that "it is wrong, always, everywhere, and for anyone to believe anything on insufficient evidence,"¹³ will give the second answer, A2, on the grounds that Abe's belief wouldn't be supported by the evidence.

A simple monist holds that for every decision situation, a single 'kind' or 'domain' of normativity matters. Here are some possible simple monist positions:

Only considerations of self-interest matter.

Only moral considerations matter.

Only moral considerations matter when the question is which action should be taken. Only evidential considerations matter when the question is which doxastic state to be in.

If simple monism is true, then we never weigh different types of considerations against each another. And there may not be different kinds of should concepts. So the simple monist doesn't have difficult or complex things to explain. The strategy can deliver one of the things we want from an ideal normative theory: a single answer about what should be done

¹³ Clifford 1877 (1886) page 346.

in each case of apparent conflict. But the cases do seem to involve some kind of relativity or conflicting normative considerations, and the simple monist strategy rules this out.¹⁴

The sophisticated monist gives one answer to each case. However, she holds that each potential answer is supported by a different kind of consideration that genuinely counts in favor of it. Her own answer is supported by the fact that one kind of consideration *outweighs* the other.

She might explain that in Bess's case there are *some* reasons in favor of both:

B1. Bess should order the beef burger.

B2. Bess should order the veggie burger.

Suppose vegetarianism is morally required. Then ordering the veggie burger is morally required, so this supports B2. But ordering the beef burger would best satisfy Bess's desires, so this supports B1. The sophisticated monist might think that moral considerations trump prudential considerations and that, taking everything into account, Bess should order the veggie burger.

Sophisticated monism is attractive because it promises everything we want: genuinely conflicting considerations but single answers about what should be done. What's needed is a story about how to weigh these considerations.

The pluralist will say that answers A1 and A2 are true in their own ways and that answers B1 and B2 are true in their own ways.¹⁵ The pluralist doesn't need to justify one answer *over*

¹⁴ Explaining away these intuitions is easier in some cases than others. For example, it may be tempting to side with a simple monist who claims that chess considerations aren't really *normative*. But this tactic is implausible when it comes to conflicting moral, prudential, and epistemic considerations.

¹⁵ Or things *like* these answers. For example, the kind of pluralist view I'll consider here is one according to which each answer must be qualified in a certain way.

another. But she does need to explain how both answers could be consistent. In the case of Carl, she might explain that both of C1 and C2:

C1. Carl should move his rook.

C2. Carl should move his bishop.

are ‘true in their own ways’ because there are two different kinds of should concepts.

There’s ‘prudential should’ to go with prudential considerations, and ‘chess should’ to go with the rules of chess. In that case C1 is true because Carl chess-should move his rook, but C2 is also true because Carl prudentially-should move is bishop.

Like sophisticated monism, pluralism promises to capture our intuition that the cases involve conflicting considerations. So it can deliver one of the things we want from an ideal normative theory. The pluralist must explain how there can be different kinds of should concepts,¹⁶ but this could be easier than explaining how to weigh different kinds of considerations against each other. The harder part, for the pluralist, is providing a theory that can tell us something satisfying about which of our options we should choose if we can only choose one. For example, while we might be willing to agree that there’s *some sense* in which Carl should play his rook and *some sense* in which he should play his bishop, we’ll probably still want to ask: “But which move *should* he make, *really?*”¹⁷ Not only that, but most of us have a pretty good idea what the right answer is.

Here’s how I think the three strategies look, at the outset, with respect to our two desiderata:

¹⁶ I mean to include views according to which there is a single should concept that comes in different ‘flavors.’

¹⁷ A pluralist like Stephen Finlay can explain that in every utterance, ‘should’ has a single, contextually-determined, meaning. (See his 2014, section 6.2.) So on his view, questions like this will always have an answer, though the explanation of what exactly ‘should’ means in these contexts is complex.

	<i>Desideratum 1: accommodating a genuine kind of relativity</i>		<i>Desideratum 2: accommodating a single/privileged answer</i>	
	Possible?	Difficulty?	Possible?	Difficulty?
Simple Monism	NO	N/A	YES	EASY
Sophisticated Monism	YES	MODERATE?	YES	MODERATE?
Pluralism	YES	EASY	MAYBE?	HARD

Table 1. How the strategies for responding to cases of apparent conflict meet the desiderata

In addressing our cases of conflict, I think the first thing to try is sophisticated monism. We can take this strategy if we can provide a good story about how to weigh normative considerations to build an all-things-considered should concept. If this can be done, then we'll have a normative theory that does both things we want. Pluralism is also worth exploring, but it's less obvious how a pluralist could deliver the second desideratum. Simple monism rules out the chance of getting the first desideratum. If neither of the other strategies can be made to work, then we might reconsider it. But I'll put it aside now, and since in the rest of the chapter I'll argue that at one of the other strategies will work, I won't return to it.

In the next section, I propose a way of thinking about normative concepts like SHOULD: the 'Normativity *Lite* Framework.' It promises to provide a satisfactory all-things-considered should concept for the sophisticated monist, or else to make pluralism acceptable. Either way, there's a solution to our cases of apparent conflict.

1.2. Normativity *Lite*

The case of Carl involves a conflict between prudence and the rules or standards of chess. For many of us, *playing a good chess game* seems different from the other kinds of considerations. In particular, it doesn't seem to *matter* in the same way that considerations like *avoiding death*, or *doing the morally right thing*, or *believing in accordance with one's evidence* do. The term 'normative authority' has been used to pick out the quality that prudence (e.g.) has and that the rules of chess (e.g.) are supposed to lack.¹⁸ If prudence and chess rules conflict, or if morality and chess rules conflict, there's somehow *no question* that chess rules lose.

On a popular way of explaining things, there are two kinds of normativity. One is 'merely formal normativity' (or 'norm normativity' or 'rule normativity'), and is generated by sets of rules or standards. Simply because there are rules of chess, there are better and worse chess moves and chess players. There's much disagreement about the nature of this kind of normativity. Perhaps it doesn't deserve to be called 'normative' at all. Or it doesn't generate reasons. Or it only 'applies' or has 'force' if the agent *wants* to follow the rules. The second kind of normativity is 'authoritative normativity' (or 'substantive normativity' or 'robust normativity'). There's also disagreement about its nature—in particular, about what makes it 'authoritative.' Furthermore, not everyone agrees that it exists. However, if it does, doing what's required by authoritative normativity has *real importance* in a way that doing what's required by some random set of rules does not.

The theory I propose can offer an explanation of the distinction between merely formal normativity and authoritative normativity. But its primary purpose is to provide a useful underlying *structure* for normative concepts like SHOULD. I'll call this underlying structure

¹⁸ For examples of this use of the term 'authoritative,' see Dorsey 2016 and Baker 2017. Other terms have also been used to make this distinction, e.g. 'substantive' as opposed to 'merely formal' normativity (see Woods (Forthcoming)), and 'true normativity' (see Broome 2013 pages 26-7).

‘normativity *lite*.’ I think most theorists will be happy to recognize it as a kind of merely formal normativity. But it matters little whether it’s actually *normative* or is just some precursor to normativity.

Lite normative concepts are defined relative to states of affairs. And some states of affairs are more valuable to us than others. (For example, *the state of affairs in which we have cured cancer in the next five years* is more valuable to us than *the state of affairs in which we have not*.) Furthermore, some states of affairs might have the kind of *real importance* or value required to give something normative authority. In the Normativity *Lite* Framework, our ordinary or authoritative concepts are just *lite* normative concepts that involve important states of affairs. Positing an ‘underlying level’ of *lite* concepts offers something that all normative theorists can agree on, even if they disagree about which states of affairs are important or most important. And, as I will argue in section 1.3, this way of thinking about things can also help us construct an all-things-considered concept that many theorists could find acceptable.

I will first present the theory as a theory of should concepts, and then describe how it could be expanded to other normative concepts.

1.2.1. The Normativity *Lite* Framework for SHOULD

Many different accounts of SHOULD (and related concepts) have been proposed. For example, where ‘ ϕ ’ stands for one of an agent ‘A’'s options:¹⁹

- i. A SHOULD ϕ *iff* ϕ ing maximizes utility.
- ii. A SHOULD ϕ *iff* by ϕ ing, A would follow the categorical imperative.
- iii. A SHOULD ϕ *iff* ϕ ing best satisfies A’s desires.

¹⁹ In what follows, I’ll assume that we have a workable definition of what an agent’s options are, and that an agent’s options are constrained by what she ‘can’ do in some salient sense of ‘can.’

However, all these accounts seem able agree on the following, so long as the state of affairs ‘S’ is correctly specified:

MIN: A SHOULD ϕ *iff* if A were to ϕ , state of affairs S would obtain.

For example:

On (i): A SHOULD ϕ *iff* if A were to ϕ , *the state of affairs in which utility is maximized* would obtain.

On (ii): A SHOULD ϕ *iff* if A were to ϕ , *the state of affairs in which A follows the categorical imperative* would obtain.

On (iii): A SHOULD ϕ *iff* if A were to ϕ , *the state of affairs in which A’s desires are best satisfied* would obtain.

On each corresponding account, the truth of right-hand side is both necessary and sufficient for it to be the case that A should ϕ .²⁰

I propose that we treat *all* subjunctive facts of the form *if A were to ϕ , then state of affairs S would obtain* as something like ‘potential’ should facts. I’ll call these ‘*lite*’ should facts. Here’s the (first pass) definition of my proposed concept ‘SHOULD_{lite}’:

SHOULD_{lite} (1): A SHOULD_{lite} ϕ (relative to S) *iff* if A were to ϕ , then S.

Should_{lite} facts are state-of-affairs-relative, so there’s no such thing as ‘what you simply should_{lite} do.’ There’s only ‘what you should_{lite} do relative to the state of affairs in which you

²⁰ Some of these accounts of should may be labeled to make it clear that they are only accounts of a particular kind of should concept. For example, (iii) might be an account of PRUDENTIAL SHOULD as opposed to SHOULD. Relativized accounts of should (e.g. cultural relativism) can also be described in a way that allows them to agree with MIN.

own a million dollars’ or ‘what you should_{lite} do relative to the state of affairs in which you’ve been invited to travel to Mars.’²¹

Here and elsewhere I intend a strict reading of the subjunctive. ‘If A were to ϕ , then S’ is true just in case S obtains in *every* possible world in which A ϕ s. A consequence of this is that some of the things I’ll go on to say are not, strictly speaking, true. But in each case I could make the appropriate adjustments; I’d just have to state things less simply.

How do we get from SHOULD_{lite} to SHOULD? Here’s the view:

The Normativity *Lite* Framework for SHOULD²²

1. There are ‘minimal’ versions of should concepts, as described by the state-of-affairs-relative concept SHOULD_{lite}.
2. States of affairs have degrees of VALUE.
3. Kinds of should concepts (if there is more than one kind), are found by locating should_{lite} concepts with valuable or otherwise notable states of affairs. *The* concept

²¹ When read aloud in a sentence, like “Alice should_{lite} turn on the TV (relative to the state of affairs in which she watches the news),” the ‘*lite*’ bit signals the state-of-affairs-relativity of the ‘should.’

²² The Normativity *Lite* Framework for SHOULD (and the more general version I present in the next section) resembles a number of pluralist and relationalist views about normativity. Some notable related views are: Stephen Finlay’s end-relationalist normative semantics (see his 2014), Evan Tiffany’s normative pluralism (see his 2007), David Copp’s teleological pluralism (see his 2009), Daan Evers’ standards-relationalism (see his 2011), and Derek Baker’s pluralism (see his Forthcoming). The four main differences between the Normativity *Lite* Framework and these other views are: (1) it distinguishes between *lite* and full-fledged normative concepts, (2) it has a narrower scope than other views, as the schema is only meant for (*lite* versions of) (option-normative) concepts, (3) it relies on a subjunctive relation to states of affairs, as opposed to, e.g., a probabilistic one or some relation to standards, and (4) it’s a revisionary as opposed to a descriptive view. I discuss Finlay’s version of relationalism in Appendix B.

SHOULD (if there is just one) will be *something like* the should_{lite} concept with the most valuable state of affairs.²³

This section has so far been devoted to introducing SHOULD_{lite} so that (1) is understandable. (2) introduces another normative concept, VALUE. I'll assume here only that our account of VALUE is *independent*: that it's not defined in terms of SHOULD_{lite}. A non-naturalist might posit an irreducible concept of value. A naturalist might hold that the value of states of affairs depends on our desires. Both accounts of value are compatible with the Normativity *Lite* Framework for SHOULD.²⁴

The framework offers two things: an extremely minimal type of should concept and a description of which tokens might be authoritative or otherwise important. These concepts would not merely be *lite* concepts in our final normative theory, and could be given special names.

However, one adjustment must be made before should_{lite} concepts can do a good job playing the role of potential should concepts. Note, first, that there's an important difference between requirements and permissions. Plausibly, if it's impossible for you to both (ϕ and ψ), then you cannot be required to ϕ *and* be required to ψ . But you could be permitted to ϕ and also be permitted to ψ . Suppose that both the high road and the low road lead home. Since you cannot take both roads, you cannot be required to take one and

²³ This isn't it, though. The most valuable state of affairs might be one like *the state of affairs in which every being that ever exists flourishes and realizes its greatest potential*. Now, nothing I could ever do would guarantee this. So if SHOULD is SHOULD_{lite} relative to *this* state affairs, then there's never anything that I should do. I identify a more promising state of affairs in section 1.3.

²⁴ I'll often speak as though not all states of affairs are equally valuable, if only because this is true. However, the account is compatible with a view of value according to which all states of affairs are equally valuable or have no value. On a view like this, there would be no difference between *lite* and other should concepts.

also be required to take the other. But you might be permitted to take one and also be permitted to take the other.

SHOULD is a ‘requirement-like’ concept: it can’t be that you both should take the high road and that you should take the low road if doing one prevents you from doing the other. But according to SHOULD_{lite} (1), you should_{lite} take the high road and you should_{lite} take the low road, since both actions would guarantee the state of affairs in which you’re home.

To ensure that SHOULD_{lite} has this same requirement-like aspect as SHOULD, I’ll adjust the definition of SHOULD_{lite} to:

SHOULD_{lite}: A SHOULD_{lite} ϕ (relative to state of affairs S) *iff* (**ϕ ing is the maximally disjunctive option such that:**) if A were to ϕ , then S.

By a ‘*maximally disjunctive* option such that if A were to ϕ then S,’ I mean one that can be picked out by a string of disjuncts such that: (i) it’s true of *each disjunct*, d , that if A were to d , then S, and (ii) A has no other option (guaranteeing S) that can be described by a longer disjunction.²⁵ On this new definition, *there’s always at most one thing that you should_{lite} do relative to any state of affairs*. In the case of the high road and the low road, you should_{lite} *take either road* relative to the state of affairs in which you’re home. This revised version of SHOULD_{lite} will play an important role in section 1.3.

²⁵ ‘Longer’ disjunctions that merely repeat disjuncts don’t count. It also only makes sense to compare disjunctions that are expressed at the same ‘level of detail.’ For example, the options *open a door* and *open a door or open a window* are both equally detailed, so the second counts as a longer disjunctive option than the first. However, the options *open a door* and *open a door with your hand or open a door with your foot* are not expressed at the same level of detail. To make them comparable, we’d need to redescribe the first option. And it might be that, once redescribed, it’s the longer disjunctive option of the two. (For example, *open a door* may be equivalent to *open a door with your hand or open a door with your foot or open a door with something else*.)

1.2.2. The General Normativity *Lite* Framework

The Normativity *Lite* Framework for SHOULD can be generalized to other related normative concepts, so long as there remains a separate concept of value for states of affairs. Given the cases of apparent conflict we began with, what we’ve wanted from SHOULD is a concept that tells us which *options* agents should perform. And from VALUE, I need a concept that ranks *states of affairs*. So I propose a partial division of normative conceptual space into normative concepts that evaluate an agent’s options (I’ll call this ‘option-normativity’), and those that evaluate states of affairs (I’ll call this ‘state-of-affairs normativity’).²⁶ I don’t claim that this division is deep or that it divides everything correctly. It’s main purpose is to locate other normative concepts that the framework can be extended to. At the very least, it separates a kind of value concept that’s needed (one that evaluates states of affairs and not options) from some other concepts that are like SHOULD.

In the general Normativity *Lite* Framework, there are ‘*lite*’ versions of all option-normative concepts. For all option-normative concepts **N**, agents A, options ϕ , and states of affairs S:

The Normativity *Lite* Schema

$A \mathbf{N}_{lite} \phi$ (relative to state of affairs S) *iff* if A were to ϕ , then S

It’s important to stress that this is only a schema, and also a rough schema at that. Just as I had to make a revision to SHOULD_{lite} to capture something important about the concept SHOULD, we’ll need more distinctions to preserve some of the finer-grained (and useful)

²⁶ Some normative theorists make a related distinction between ‘deliberative’ and ‘evaluative’ normativity. See, for example, Schroeder 2011.

differences between our current normative concepts. For example, by directly applying the schema, we'd get *lite* versions of common option-normative concepts like:

A SHOULD _{lite} ϕ (relative to state of affairs S)	<i>iff</i>	if A were to ϕ , then S
It's PERMISSIBLE _{lite} for A to ϕ (relative to state of affairs S)	<i>iff</i>	if A were to ϕ , then S
A HAS A REASON TO _{lite} ϕ (relative to state of affairs S)	<i>iff</i>	if A were to ϕ , then S
It's RATIONAL _{lite} for A to ϕ (relative to state of affairs S)	<i>iff</i>	if A were to ϕ , then S
A is BLAMELESS _{lite} for ϕ ing (relative to state of affairs S)	<i>iff</i>	if A were to ϕ , then S

I've said that this is a *potentially revisionary* project, but collapsing all our option-normative concepts into one isn't progress. It's beyond the scope of this chapter to survey all the potential revisions we could make, but I discuss some possibilities in Appendix A.

Once we make these distinctions, the more general Normativity *Lite* Framework is:

The Normativity *Lite* Framework

1. There are 'minimal' versions of option-normative concepts, roughly described by the Normativity *Lite* Schema.
2. States of affairs have degrees of VALUE.
3. Ordinary option-normative concepts are found by locating *lite* concepts with valuable or otherwise notable states of affairs.

In the Normativity *Lite* Framework, the relevant concept of value won't fit the Normativity *Lite* Schema. However it works, its job is to rank states of affairs and it's the main thing that determines which, if any, kinds of *lite* option-normative concepts are authoritative. If we like, we can also choose to recognize and name other kinds of option-normative concepts.

Some of these (e.g. ‘CHESS SHOULD’ and ‘CHESS REASONS’) could be merely formally normative. Others, like ‘EPISTEMIC REASONS’ might be somewhat valuable and so have some authority, but may not be the most authoritative.

A natural question to ask in this framework is: Is there a *most* authoritative or *most* important kind of *lite* option-normative concept?

1.3. An All-Things-Considered SHOULD

The Normativity *Lite* Framework is compatible with both sophisticated monism and pluralism. Here’s how to build a plausible all-things-considered SHOULD concept for the sophisticated monist.

First, assume the revised account of SHOULD_{lite} I’ve proposed:

A SHOULD_{lite} ϕ (relative to state of affairs S) *iff* **(ϕ ing is the maximally disjunctive option such that:)** if A were to ϕ , then S

And recall that ‘ ϕ ’ always stands for one of A’s options. Whenever you have options, many ‘SHOULD_{lite} facts’ are true of you. For example: *you should_{lite} check a clock relative to the state of affairs in which you know the time, you should_{lite} wear a fake Hitler mustache relative to the state of affairs in which you learn what that’s like, and so on.* This set of SHOULD_{lite} facts includes a large number of states of affairs.²⁷

²⁷ There are, however, states of affairs that won’t appear in this set. For example, consider the state of affairs in which *you own a square circle*, or the one in which *you solve an unsolvable problem*. There will never be true SHOULD_{lite} facts involving you and them, since you’ll never be able to do anything that would guarantee them.

Call states of affairs that appear in true $\text{SHOULD}_{\text{lite}}$ facts about you ‘*reachable*.’ These are all the states of affairs that you can bring about by performing one of your options. And assume that your reachable states of affairs can be given a single ordinal ranking according to their value. Then there will be a set of your *most valuable reachable states of affairs* containing only reachable states of affairs that are no less valuable than any others.

Now consider the state of affairs I’ll call ‘ S^* : *doing what you should_{lite} do relative to one of your most valuable reachable state of affairs*’. If you succeed in bringing about this state of affairs, then you’ll have brought about one that’s at least as valuable as any other state of affairs that you could have brought about.

Suppose I’m camping. The $\text{SHOULD}_{\text{lite}}$ -facts about me are:

I should_{lite} gather firewood relative to the state of affairs in which I stay warm at night.

I should_{lite} repair my tent relative to the state of affairs in which I stay dry at night.

I should_{lite} gather firewood and repair my tent relative to the state of affairs in which I stay warm and dry at night.

I should_{lite} scatter food around my campsite relative to the state of affairs in which I attract bears.

My reachable states of affairs are, from most to least valuable:

1. *staying warm and dry at night*
2. *staying dry at night*
3. *staying warm at night*
4. *attracting bears*

Since my most valuable reachable state of affairs is *staying warm and dry at night*, I should_{lite} gather firewood and repair my tent relative to S^* . If I do both chores, I’ll ensure that the most valuable state of affairs obtains, given the options I have.

The concept *should_{lite} relative to S** is an all-things-considered one. It takes into consideration every option one has, every possible consequence of performing them, and the amount of value that each of these ‘consequences’ has.²⁸ With it we can define a full-fledged, ‘all-things-considered’ SHOULD concept:

A (ALL-THINGS-CONSIDERED) SHOULD ϕ iff if A were to ϕ , then *S** (*A will have done what A should_{lite} do relative to one of A’s most valuable reachable states of affairs*)

A all-things-considered should make pancakes is equivalent to *A should_{lite} make pancakes relative to S**, the state of affairs in which A has done what she should_{lite} do relative to one of A’s most valuable reachable states of affairs. So this full should concept is identical to a *lite* should concept. That’s how the framework works: the *lite* concepts populate normative conceptual space with every possible (option-)normative concept we could need, and then it’s up to us to pick out and re-name the important ones.

In the Normativity *Lite* Framework, it’s possible to give sophisticated monist responses like the following to the three cases of apparent conflict I started with:

Abe should_{lite} not believe that he’ll recover (relative to *the state of affairs in which he only believes propositions likely to be true*). He should_{lite} believe that he’ll recover (relative to *the state of affairs in which he maximizes his chance of recovery*). And he all-things-considered should believe that he’ll recover because the second state of affairs is more valuable than the first.

²⁸ This talk of consequences isn’t quite right, because the theory isn’t a version of consequentialism. What the concept *should_{lite} relative to S** really takes into consideration are all of the subjunctive truths about what would be the case conditional on your performing each of your options. Many of the things that would be the case if you were to perform one of your options will not be consequences of them at all, and some may only be consequences in the most strained sense (e.g. a ‘consequence’ of my jumping is that I’m jumping).

Bess should_{lite} order the veggie burger (relative to *the state of affairs in which she maximizes utility*). She should_{lite} order the beef burger (relative to *the state of affairs in which she satisfies her desires*). And she all-things-considered should order the veggie burger because the first state of affairs is more valuable than the second.

Carl should_{lite} play his rook (relative to *the state of affairs in which he wins the game*). He should_{lite} play his bishop (relative to *the state of affairs in which he stays alive*). And he all-things-considered should play his bishop because the second state of affairs is more valuable than the first.

1.4. An Adequate Version of Pluralism

Back in section 1.1 when I introduced pluralism, I said that the pluralist must explain how two seemingly inconsistent answers can be true in cases of apparent conflict. I also said that pluralism *might* be able to offer privileged answers to questions about what should be done in these cases. Let's begin with the first task. The Normativity *Lite* Framework provides a few different ways of doing it.

The most obvious is to note that there will be at least two apparently conflicting answers to questions about what *should_{lite}* be done in each case. The pluralist could say some of the same things as the sophisticated monist in my example above:

Abe should_{lite} not believe that he'll recover (relative to *the state of affairs in which he only believes propositions likely to be true*). He should_{lite} believe that he'll recover (relative to *the state of affairs in which he maximizes his chance of recovery*).

Bess should_{lite} order the veggie burger (relative to *the state of affairs in which she maximizes utility*). She should_{lite} order the beef burger (relative to *the state of affairs in which she satisfies her desires*).

Carl should_{lite} play his rook (relative to *the state of affairs in which he wins the game*). He should_{lite} play his bishop (relative to *the state of affairs in which he stays alive*).

But the pluralist doesn't have to stick to *lite* concepts. She can also claim that there are full-fledged domain-relative kinds of should and that these also offer merely apparently conflicting answers to questions about what should be done. Here are some examples of domain-relative kinds of should:

A (PRUDENTIALY) SHOULD ϕ iff if A were to ϕ , then *the state of affairs in which A's desired are satisfied* would obtain.

A (MORALLY) SHOULD ϕ iff if A were to ϕ , then *the state of affairs in which A has maximized expected utility* would obtain.

A (EPISTEMICALLY) SHOULD ϕ iff if A were to ϕ , then *the state of affairs in which A's beliefs are likely to be true* would obtain.

A (CHESS) SHOULD ϕ iff if A were to ϕ , then *the state of affairs in which A plays a best chess move* would obtain.

So the pluralist can be a pluralist about SHOULD, not just SHOULD_{lite}. She'd explain that each of these states of affairs is notable in some way that warrants defining special full-fledged kinds of SHOULD connected with them. That would allow her to give the following pairs of consistent answers to the cases:

Abe epistemically should not believe that he'll recover. But he prudentially should believe that he'll recover.

Bess morally should order the veggie burger. But she prudentially should order the beef burger.

Carl chess should play his rook. But he prudentially should play his bishop.

The sophisticated monist can't posit these domain-relative SHOULDs. She'd claim that the ALL-THINGS-CONSIDERED SHOULD defined in terms of S^* is the only kind of SHOULD.

But is this the only difference between the two views? Now that we can make more distinctions, I'll define each view more carefully and distinguish between two kinds of pluralism:

Sophisticated monism. One, and only one, of an agent's options is what she ALL-THINGS-CONSIDERED SHOULD do. There may be other things that the agent should_{lite} do, but this is the only SHOULD concept, and it has a greater degree of authority than any other choice.

Weak Pluralism. It's possible that an agent should_{lite} do one of her options and also should_{lite} do a different one instead. And it's possible that an agent should do one of her options and also should do a different one instead, for appropriate kinds of SHOULD.

(Strong) Pluralism. It's possible that an agent should_{lite} do one of her options and also should_{lite} do a different one instead. And it's possible that an agent should do one of her options and also should do a different one instead, for appropriate kinds of

SHOULD. Furthermore, there is no single ALL-THINGS-CONSIDERED SHOULD concept that has a greater degree of authority than all other other SHOULD concepts (or SHOULD_{lite} concepts).

In the Normativity *Lite* Framework, the sophisticated monist and weak pluralist both agree that there can be different *lite* should concepts, but the monist thinks there can be only one should concept. While this is the kind of disagreement that gives the two views their names, in the new framework it doesn't matter all that much. Furthermore, if there's any good reason to call other *lite* concepts by special names (e.g. because it's useful), then the weak pluralist seems to have the better position here.

It's the disagreement between the sophisticated monist (or anyone who thinks there can be a single most authoritative all-things-considered should concept) and the strong pluralist that's most important in the Normativity *Lite* Framework.

The sophisticated monist is committed to the existence of an all-things-considered kind of SHOULD that's more authoritative than any other kind of SHOULD or SHOULD_{lite}. And the all-things-considered concept I proposed in the previous section does seem more authoritative than any other choice.²⁹

So, does this show that sophisticated monism is true and that strong pluralism is false? Unfortunately, no. The strong pluralist can agree that *if* my all-things-considered should concept is available, then sophisticated monism could be true. However, strong pluralists don't think the concept makes any sense. Why? Everything I used to define it fell out of the

²⁹ And as far as I can tell, it's the only possible all-things-considered should concept that could work for the sophisticated monist in the Normativity *Lite* Framework. However, strictly speaking, all the sophisticated monist needs is an all-things-considered kind of SHOULD that's at least as authoritative as any other kind of SHOULD, and that never disagrees with an equally-authoritative kind of SHOULD.

Normativity *Lite* Framework, except for one crucial thing: *the assumption that there's a single privileged ordinal ranking of states of affairs according to their values.*

This is the only thing that it's plausible for the strong pluralist to deny, and she does deny it. In the Normativity *Lite* Framework, it's the core of the disagreement between the sophisticated monist and the pluralist. (From now on 'pluralist' will just mean 'strong pluralist' unless otherwise specified.) If there is a single privileged ranking, then we should adopt the sophisticated monist's strategy for cases of apparent conflict. But if there is not, then (strong) pluralism will be an acceptable strategy.

Before turning to that, the last thing to do in this section is to consider whether a pluralist could deliver the second desiderata for normative theories we began with: Can the pluralist give us a single privileged answer to should-questions in cases of apparent conflict? Both kinds of pluralists can follow Stephen Finlay and adopt a contextual semantics for should-questions.³⁰ On this view, any question like 'what should Carl do, really?' picks out a single should or should_{lite} concept and can be given a single answer to. However, then the important issue becomes: what sort of should concepts are available? This would not be a very satisfying move if the only should concepts available in Carl's case are CHESS SHOULD and PRUDENTIAL SHOULD and the question of *what Carl should do, really* must use one of these. What we'd want is for there to be *some* kind of all-things-considered concept available.

The weak pluralist can posit the existence of the same ALL-THINGS-CONSIDERED SHOULD that the sophisticated monist uses, so weak pluralism can do just as good a job as sophisticated monism at delivering the two desiderata. But the strong pluralist can't posit a single most authoritative should concept like this. She could posit weaker 'all-things-considered' concepts that give some of the privileged answers to should-questions that we

³⁰ See, e.g., Finlay 2014 section 6.2.

want. (Like a ‘taking both chess and prudential considerations into account’ kind of should.) But the extent to which she can do this, and to which it is reasonable to do this, will depend on *why* she rejects a single privileged all-things-considered concept.

1.5. Monism *versus* Pluralism

Why might a (strong) pluralist deny the existence of a single privileged ordinal ranking of states of affairs according to their value? Here are some possible explanations:³¹

1. She thinks that there’s only **domain-relative** VALUE, and so only equally-authoritative domain-relative rankings and ALL-THINGS-CONSIDERED SHOULDs.
2. She is some other common kind of relativist about value; for example, a **cultural relativist** or a **subjectivist**. If so, there wouldn’t be any privileged ranking.
3. She has a **teleological** account of value, according to which there is no final value: states of affairs have value only relative to, and in so far as they’re conducive to, other states of affairs.
4. She denies the **existence** of VALUE, and so the existence of rankings of states of affairs according to their value.
5. She thinks that any plausible ranking of states of affairs by their values will be **intransitive**, so that in some cases there won’t be a ‘most valuable’ reachable state of affairs.

³¹ This list may not be exhaustive.

6. She thinks that some pairs of states of affairs are **incommensurable**, so that there's no fact of the matter about which is most valuable.
7. She thinks that any plausible ranking of states of affairs by their values will have **no upper bound**, so that in some cases there won't be a 'most valuable' reachable state of affairs.

The first three views are kinds of relativism about value. The first is a natural one to hold. If you're already willing to posit different normative domains that concern different kinds of things and that generate different domain-relative should facts, then why not think that they also generate different domain-relative value facts? There could also be all-things-considered value facts about which states of affairs are all-things-considered most valuable. But then we'd need an explanation of how these different considerations are weighed against each other. And I've appealed to facts about value to do this for option-normative concepts, so the same strategy wouldn't be available for constructing a concept of all-things-considered value.

The second kind of view is only as plausible as cultural relativism or subjectivism. Since these aren't especially popular views, I'll put them aside. The third view is much more compelling once you've accepted the Normativity *Lite* Framework. Within it, option-normativity is given a broadly teleological account. So it makes more sense to try to fit other kinds of normativity into a similar mold. However, the pluralist owes us an argument for this view. As she does if she holds the fourth, nihilist, view, which is the least plausible of the bunch.

The fifth, sixth, and seventh views are not relativist views. They don't claim that there's more than one way to generate the ranking (at least not at first). The worry is that the single

privileged ranking we're hoping for won't have the right kind of *structure* to do what the sophisticated monist wants. This kind of worry is a smaller one in that it may only effect certain decision scenarios. For example, even if there is no upper bound of value, this need not effect any of our actual decisions. Likewise, intransitive³² rankings of states of affairs and incommensurability³³ may only matter to some decision problems. Nevertheless, any of these or related phenomena would still be problematic for sophisticated monism, and indeed for any theory of SHOULD or other option-normative concept.

Here's how things seem to stand. 2, 3, and 4 are not of immediate concern to the sophisticated monist. They require more argument, and without it they aren't particularly plausible. 5, 6, and 7 should concern the sophisticated monist, but they also require more argument.³⁴

It's 1, the view that there are domain-relative *values* and value-rankings, that should make us most nervous about sophisticated monism and the possibility of a single authoritative ALL-THINGS-CONSIDERED SHOULD. And it would be troubling if there's only domain-relative value. The pluralist could offer us a collection of domain-relative 'all'-things-considered SHOULDs, but in many cases we wouldn't have consensus about what should be done.

But my aim here is not to defend my account of ALL-THINGS-CONSIDERED SHOULD against this kind of worry. Instead, it's to argue that the Normativity *Lite* Framework can offer *a* successful strategy for accommodating cases of apparent conflict. If there is a single

³² See, e.g., Temkin 2012 chapters 6, 7, and 13.

³³ See, e.g., Chang 1997 and 2007.

³⁴ Even better, they *might* be the sorts of problems that we could ultimately solve by building an 'all'-things-considered value ranking that considers *even more things*. For example, it might involve *meta*-values that meta-value everything that's value-incommensurable or in an value-intransitive loop transitively, either equally or by some complicated function. And the 'no-upper-bound' problem might be addressed by claiming that in such decision scenarios, there's nothing that all-things-considered should be done.

privileged way of ranking states of affairs according to their values, the strategy will be sophisticated monism. If there's no such ranking, the strategy will be pluralism. Since it will be impossible to offer a privileged all-things-considered SHOULD concept, pluralism will be the most acceptable strategy.

In this chapter, I have described a new way of organizing normative conceptual space: the Normativity *Lite* Framework. I claim that it can offer the best available strategy for accommodating cases of apparent conflict. It also sheds light on the debate between normative monists and pluralists and offers a novel account of all-things-considered normative concepts.

2

The Incompleteness Problem for Normativity

In this chapter I present an argument for the claim that it's impossible to have a complete normative theory. According to the 'Incompleteness Argument,' a complete normative theory would be able to answer every normative question. But for any theory, we can find questions of the form *what should one do, given one's uncertainty about such-and-such?* that the theory can't answer. So no normative theory could be complete. I make the case that the Incompleteness Argument, or something close enough to be concerning, succeeds.

2.1. The Incompleteness Argument

Consider the following argument:

The Incompleteness Argument

1. A complete normative theory would be able to answer every normative question.
2. For any normative theory, N, there's a question of the form *what should one do given one's uncertainty about N?* that N can't answer.
3. Therefore, no normative theory could be complete.

In this section I'll explain the argument and give reasons for thinking it succeeds. In section 2.2 I'll discuss some objections.

Premise 1 is my gloss on ‘complete normative theory.’ “What’s a *complete* normative theory?” “It’s one that could answer every normative question.” This is not intended to be controversial. For example, a complete theory of gravity might be expected to answer every ‘gravity question.’ If, say, we want to know what the force of gravity is on some alien planet, we expect our theory of gravity to be able to give the answer—assuming that we provide enough other information (like facts about the mass and density of the planet).

Premise 2 does all the work in the Incompleteness Argument, and so needs the most explanation and support. I’ll begin by assuming that a ‘normative theory’ is something that can be written in the form *A should ϕ iff ____*, where ‘A’ is an agent, ‘ ϕ ’ is an option, and the ‘____’ is filled in by some set of conditions. For example, the blank might be filled in by *ϕ ing maximizes utility*, by *ϕ ing is what a virtuous person would do*, by *ϕ ing is a state of believing that’s a product of a reliable process*, etc. This assumption restricts our scope in two ways. We ignore normative theories that are not about what *agents should do*. And we also ignore versions of pluralism or relativism about *should*, which can’t be expressed in this form. In section 2.2, I consider whether lifting these restrictions could help. But simplifying matters here allows me to give a more straightforward explanation of the argument.

Suppose our normative theory is:

Utilitarianism *A should ϕ iff ϕ ing maximizes utility*

It might seem that ‘uncertainty about the normative theory *Utilitarianism*’ just means uncertainty about its truth. But I will mean something broader than that here. You are uncertain about the normative theory *Utilitarianism* just in case you’re either:

- a. uncertain whether it's true that *A should ϕ iff ϕ ing maximizes utility*, or
- b. uncertain, for some specific option, whether performing it maximizes utility (i.e. whether the right-hand side of the biconditional is true for some instance of ϕ).

We can call the first ‘truth uncertainty’ about *Utilitarianism*, since it involves being uncertain about the truth of a normative theory. We can call the second ‘application uncertainty’ about *Utilitarianism*, since it is uncertainty about how the normative theory applies to some case or what the theory says about it.³⁵ It’s possible to have each kind of uncertainty separately or together.

What’s meant by a question like *what should one do, given one’s uncertainty about N?* The only thing that’s obvious so far is that such a question presupposes that one is uncertain about the normative theory N.

To help explain what’s meant, consider the following case:

Parfit’s Miners Case

Suppose that ten miners are trapped, and that they’re either all in mineshaft A or all in mineshaft B. Floodwaters are rising in both shafts and Tanya must decide which floodgate to close before she can find out where the miners are. If she blocks shaft A and the miners are in shaft A, then all ten will be saved. Otherwise, all ten will die. If she blocks shaft B and the miners are in shaft B, then all ten will be saved. Otherwise, all ten will die. If Tanya blocks neither shaft, allowing the water to partially flood each

³⁵ The first kind of uncertainty has been given the term ‘normative uncertainty’ elsewhere, and the second has been given the term ‘descriptive uncertainty.’ (E.g. MacAskill 2014.) I’ve chosen different terminology here because it is more accurate (at least for my purposes), and because it avoids potentially confusing phrases like ‘normative uncertainty about normative theories.’

one, then nine miners will be saved and one will die, regardless of which shaft the miners are in.³⁶

		Possible states	
		<i>Miners are in Shaft A</i>	<i>Miners are in Shaft B</i>
Tanya's options	<i>Block shaft A</i>	10 live	0 live
	<i>Block shaft B</i>	0 live	10 live
	<i>Block neither shaft</i>	9 live	9 live

Table 2. Decision table for Parfit's Miners Case

We can take two different perspectives on cases like this. 'Subjectivists' think that facts about what one should do are sensitive, in some relevant way, to one's mental states. Since Tanya has no idea which shaft the miners are in, the subjectivist's verdict about what she should do won't depend on where the miners are. For example, a subjectivist may react to the Miners Case by claiming that Tanya should block neither shaft, thereby guaranteeing that nine miners are saved. An alternative subjectivist verdict is that Tanya should choose at random between blocking shaft A and blocking shaft B, thereby doing her best to increase the likelihood that no miner dies. What these responses share in common is that the answer is given, or can be given, in a way that doesn't rely on the truth of the matter about where the miners are.

In contrast, 'objectivists' think that facts about what people should do aren't constrained in this way. For an objectivist, it doesn't matter whether Tanya knows where the miners are or not. Suppose that they're actually in shaft A. Then Tanya should block shaft A, according to the objectivist, since this act would in fact lead to the best outcome. No plausible objectivist theory gives a different verdict about what Tanya should do. (Unless we

³⁶ Parfit 1988 pages 2-3. I've rephrased things a bit. Parfit credits Donald Regan for first posing a case of this kind in Regan 1980 page 265.

add in further facts to the case, like that the miners are actually a group of terrorists with plans to release a deadly virus later that day.) But it's not hard to come up with theories that would give a different verdict while still counting as objectivist. For example, a theory that says that we should always act so as to *minimize* utility would hold that Tanya should block the shaft that the miners *aren't* in, ensuring that they'd all die. I note these alternatives to show that, in considering whether a theory is subjectivist or objectivist, it's not the answer given that matters, but rather the explanation behind it.

Subjectivists and objectivists may have a real, substantive, disagreement here about what should be done. Or, it might be resolved by replacing the single notion *should* with a subjectivist variety and an objectivist variety. But however the disagreement is understood, it's reasonable to conclude that there are two different perspectives one can take about cases like Parfit's, and that it makes sense for Tanya to do something different according to each perspective.

I'll express this here as the claim that *we can ask two different kinds of normative questions* about cases like this. These questions might be phrased in a number of ways, but I'll go with:

What should one do, objectively speaking? and,

What should one do, given one's uncertainty about such-and-such?

The first, 'objectivist question,' requires an answer that doesn't give one's state of uncertainty any special weight; the second, 'subjectivist question,' requires an answer that does.

How does this work, exactly? I'll say a bit more about the subjectivist question in section 2.2.3. But it's difficult to be too precise without making controversial claims about

how to define subjectivism and objectivism. What I will point out here is that if you are an objectivist talking to a subjectivist, and all you want is the answer to the question you care about, you can expect to get this answer by asking something like *okay, but what should Tanya do, objectively speaking?* The subjectivist could claim that you really mean to ask a different question like *what would it be **best** for Tanya to do?* But that's fine; you'll still get the answer you were after. And if you're a subjectivist talking to an objectivist, you'll usually be able to get the answer to the question you care about by asking something like *what should Tanya do, given her uncertainty about where the miners are?* The objectivist may claim that you really mean to ask a question like *what would Tanya be **blameless** for doing?* But again, you'll still get the answer you were looking for.

An answer to the question *what should one do given one's uncertainty about N?* must be sensitive to one's uncertainty about the normative theory N, just as a subjectivist's verdict about what Tanya should do in the Miners Case is sensitive to Tanya's uncertainty about where the miners are. What's meant by a theory's being 'sensitive to one's uncertainty about something'? My gloss is that the theory *does not rely on the truth about that something in giving its answer*.

Here ends my explanation of what Premise 2 means. Is the premise true? Is it true that no normative theory, N, could answer a question of the form *what should so-and-so do given her uncertainty about N?* Recall my token objectivist normative theory:³⁷

Utilitarianism *One should ϕ iff ϕ ing maximizes utility*

³⁷ I appeal to versions of consequentialism in all the examples I give in this chapter. This is merely for convenience. The Incompleteness Argument is meant to apply to any kind of normative theory.

It's easy to suppose that Tanya is not uncertain about *Utilitarianism*, since she may not have considered it before. But we can imagine a version of the case in which she does consider it, and in which she's certain that it's true.³⁸ In a case like this, she would still have application uncertainty about it. Because she doesn't know where the miners are, she wouldn't know which of *blocking shaft A* and *blocking shaft B* would maximize utility.

Suppose that's what's going on. Then, according to *Utilitarianism*, Tanya should block the shaft the miners are actually in (say, shaft A). But this is not the sort of answer that's required by a question like *what should Tanya do, given her uncertainty about Utilitarianism?* Because it depends on the actual facts about where the miners are, and because Tanya's uncertainty about these facts is exactly why she's uncertain about *Utilitarianism*, it's not appropriately sensitive to Tanya's uncertainty.

This version of the Miners Case shows that one theory can't answer a question of the form *what should one do, given one's uncertainty about [that theory itself]*? In it, Tanya has application uncertainty about a normative theory. What about cases involving truth uncertainty about a normative theory? Consider a similar kind of case, adapted from MacAskill 2014:³⁹

MacAskill's Chimpanzee Case

Susan is a doctor, and sees that patients A, B, and C all suffer from the same illness. A is human. B and C are chimpanzees. Susan has one vial of a drug, D. If she gives it all to A, A will be completely cured (A's health = 100%). If she gives it all to B and C (dividing it between them), both will be completely cured (B's health = 100% and C's health = 100%).

If she splits the drug between all three, A will be almost cured (A's health = 99%) and B and

³⁸ Or a case in which she's certain enough to assume its truth for the sake of her decision. See Appendix C for a discussion of what I mean by 'uncertainty.'

³⁹ MacAskill 2014 pages 9-10. I've changed the case slightly to make Susan's truth uncertainty explicitly about two different moral theories.

C will be partly cured (B's health = 50% and C's health = 50%). Finally, suppose that Susan's confidence is divided equally between two moral theories. According to *M1*, *one should ϕ iff ϕ ing maximizes moral value^H*, where only human wellbeing contributes to moral value^H. According to theory *M2*, *one should ϕ iff ϕ ing maximizes moral value^{HC}*, where the wellbeing of humans and chimpanzees contribute equally to moral value^{HC}.

		Possible states	
		<i>M1 is true (chimps' health has no value)</i>	<i>M2 is true (chimps' and humans' health have equal value)</i>
Susan's options	(1) Give D to A	100	100
	(2) Split D three ways	99	199
	(3) Give D to B + C	0	200

Table 3. Decision table for MacAskill's Chimpanzee Case

Our protagonist, Susan, has truth uncertainty about which of *M1* and *M2* is true. If we suppose that *M1* is true, then, objectively speaking, Susan should give the drug to A since that would maximize moral value according to *M1*. (It would maximize moral value^H.) But what should Susan do, given her uncertainty about *M1*? *M1* can't answer the question *what should Susan do, given her uncertainty about M1?* because it's insensitive to Susan's state of uncertainty. *M1*'s answer presupposes something that Susan is uncertain about: that *M1* is true. So we have a case involving truth uncertainty in which a theory can't answer a question of the form *what should one do, given one's uncertainty about [that theory itself]?*

What would an answer to the question *what should Susan do, given her uncertainty about M1?* look like? It seems to many that, given her uncertainty about *M1*, Susan should split the

drug three ways.⁴⁰ Because option 2 (splitting the drug three ways) does very well no matter which normative theory is true, one needn't rely on the truth of either theory in giving this answer.

I've just described two normative theories that can't answer a question of the form *what should one do given one's uncertainty about [the normative theory]*? Is it true that *no* normative theory, N, could answer a question of the form *what should one do given one's uncertainty about N*? It seems that way.

For a normative theory of the form *one should ϕ iff p* to be sensitive to someone's application uncertainty about it, it would have to fail to rely on the truth about whether or not p is true. And given how normative theories work, this is impossible. The theory will 'say' that "one should ϕ " if, and only if, p . Imagine for a moment that each theory is actually a person who is certain of the theory and has access to any facts she needs in order to answer normative questions about what agents should do. Then this person must always be able to help herself to the truth about whether or not p obtains in order to provide her answers. For example, if you were to ask the *Utilitarianism*-person what Tanya should do, she would need to help herself to the actual location of the miners in order to respond.

What about truth uncertainty? Could a normative theory of the form *one should ϕ iff p* be sensitive to someone's uncertainty about whether it's true? This would be no easier. To be sensitive to someone's truth uncertainty about it, the theory would have to fail to assume its

⁴⁰ I'll note that there's plenty of disagreement about truth uncertainty (also known as 'normative uncertainty,' or, if about moral theories, 'moral uncertainty'). I assume here that the intuitive judgment in these cases is that: there's at least one important sense in which someone who is uncertain about whether a normative theory is true may not be required to act in accordance with it. The opposing position is that even if there's an important sense in which Tanya should take the less risky option and choose to block neither shaft in the Miners Case, parallel reasoning doesn't apply in cases of truth uncertainty. Defenders of this alternative view include Weatherson 2014, Harman 2015, Hedden 2015, and Nissan-Rozen 2015.

own truth in issuing verdicts. But again, using the analogy of the theory as a person, it seems that the person must always be able to help herself to the truth of the theory in order to provide the answers to questions about what agents should do. She'd interpret them as being equivalent to questions about what they should do, *according to the theory*, which is a give-away that she's going to assume the truth of the theory in answering. For example, if you were to ask the 'maximize moral value'^H theory (*M1*) what Susan should do, the *M1*-person would first assume that *one should ϕ iff ϕ ing maximizes moral value^H*, then check to see which option would maximize moral value^H, and then answer that Susan should perform this option.⁴¹

Premise 2 says that: 'For any normative theory, *N*, there is a question of the form *what should one do given one's uncertainty about N?*' So far, I've argued that no normative theory could answer a question of this form. But could a question of this form be asked of any normative theory? The question presupposes that one is uncertain about the theory. So if there's a normative theory that it's impossible to be uncertain about, then this premise could be false.

However, it is possible to be uncertain about any normative theory. Let's distinguish between trivially true and substantive normative theories. If a theory is trivially true, then it may be difficult to doubt its truth. Consider, for example, the very simple trivial theory (or 'theory'): *one should ϕ iff one should ϕ* . It may be impossible to doubt this, at least if you understand it. However, if you're uncertain about what you should do, then you're going to have application uncertainty about this theory. And this counts as a way of 'being uncertain

⁴¹ Note that a normative theory's inability to be sensitive to uncertainty about itself does not mean that it must be 'immodest.' In the literature on disagreement, for example, the 'conciliationist' theory is sometimes said to be modest, because if one is uncertain between it and a 'steadfast' theory, then it might advise you to lower your confidence in itself. (See, e.g., Christensen 2013.) However, whatever it might advise *you* to do, it does this by relying on its own truth. If we think of theories as perfect reasoners who must 'answer' questions by providing proofs with the answers as their conclusions, then we can look to the lines in the proof to see what has been assumed.

about it.’ So trivial theories would only be immune to uncertainty if it were also impossible to be uncertain about what we should do. Since we’re obviously often uncertain about what we should do, it’s easy to imagine someone who’s uncertain about any trivial normative theory. Moreover, if trivial theories were the only ones to escape the Incompleteness Argument, this wouldn’t be much cause for celebration.

What about substantive normative theories? Because they are ‘substantive,’ it will be possible to doubt their truth. And that’s all that’s needed for the claim that it’s possible to be uncertain about them.

It may also be possible to have application uncertainty about any normative theory. It’s uncontroversial that claims about the external world can be doubted. But many of us assume that our own mental states are immune from doubt. If asked something like “Are you *sure* you turned off the stove?” we often assume that we can safely respond “Yes, I’m sure I turned it off.” We don’t expect to be pressed further with “But are you *sure* that you’re sure you turned it off?”

Subjectivist normative theories often look like they’re designed to be responsive to the first kind of question but not the second. For example, we don’t always know which of our actions will maximize utility. And it’s thought to be a shortcoming of objective utility theory that it can’t answer all our questions about what we’re to do when we’re not sure what will maximize utility. Enter *expected* utility theory. Many theorists seem to hope that replacing an objectivist normative theory with a *more* subjectivist one solves the problem. But this only solves the immediate problem. It could only work completely if the more subjective notion the theory depends on (expected utility as opposed to utility) is one that it’s impossible to be uncertain about.

Timothy Williamson makes the case that no non-trivial mental state is immune to doubt in his (2000) and (2008). In one of his famous ‘anti-luminosity arguments,’ he argues that even mental states like *feeling cold* might be such that you can be in them without realizing that you are. His argument proceeds by supposing that at time 1 you’re both cold and aware that you’re cold and that at some later time 2 you’re both warm and aware that you’re warm. Then he has us imagine that the temperature increases gradually between time 1 and time 2. What he tries to show is that there’s a certain point where you still feel cold, but are no longer in a position to know it.⁴² In the context of this chapter, we can say that at this point you feel cold but could be uncertain about whether you do.

Whether or not Williamson’s anti-luminosity arguments succeed is up for debate. But if they do, then it may be possible to have application uncertainty about even the most subjective normative theory. And there’s another way to motivate the claim that it’s possible to be uncertain about your mental states. I’ll simply ask you to produce a correct definition of the mental state in question. For example, consider ‘believing,’ ‘wanting,’ and ‘seeming.’ It’s not easy to produce definitions that you can be certain about. There may be good alternative definitions, and you may be able to imagine some borderline or hard-to-classify cases.

For example, I’m disposed to react with terror when I encounter a spider. But I can also tell you, calmly and convincingly, that I know that most spiders do not pose any real danger. Do I count as believing that spiders are dangerous, or not?⁴³ Or, suppose that I salivate when offered a piece of cake, but tell you I don’t want it. Am I telling the truth? Finally, suppose I’m given a blind taste test between Coke and Pepsi. Couldn’t it be accurate for me

⁴² Williamson 2000 pages 96-7.

⁴³ For a discussion of this issue see, e.g., Schwitzgebel 2010.

to say: *They're so similar. The first might seem a little sweeter. Or maybe it seems a bit tangier? I'm not really sure.* If you can imagine cases like these, then you can imagine cases in which someone is uncertain about her own mental states.

We now have compelling support for the Incompleteness Argument. It's possible to be uncertain about any normative theory. (Either about its truth or about how it applies to some case.) So for any normative theory, N, there's a question of the form *what should one do, given one's uncertainty about N?* And no normative theory, N, can answer a question of that form. Therefore, Premise 2 is true: for any normative theory, N, there's a question of the form *what should one do, given one's uncertainty about N?* that N can't answer. Finally, since by Premise 1 a complete normative theory is one that could answer every normative question, it follows that no normative theory could be complete.

2.2. Objections and Ways Out

In section 2.1 I introduced the Incompleteness Argument. In the rest of the chapter I will consider attempts to avoid it and argue that none succeed. The idea that a normative theory can have trouble accommodating cases of uncertainty about itself is not new. But my aim is to show that the problem is much more difficult than it has been taken to be. I'll distinguish between three kinds of responses to the Incompleteness Argument. You might take up an *optimistic* attitude, and approach it as challenge that can be solved with enough ingenuity. You might take up a *fastidious* attitude, scrutinizing the premises to try to show that the argument is unsound. Or, you might take up a *dismissive* attitude, and argue that completeness is unimportant or that complete normative theories don't need to do what the

Incompleteness Argument demands of them. I'll give examples of each kind of response to show how unlikely it is that there's a satisfying way out.

2.2.1. 'Optimistic' Responses

2.2.1.1. Hierarchy Theories

In presenting the Incompleteness Argument, I made the assumption that normative theories could be expressed in the form *one should ϕ iff _____*. But there are more complicated kinds of theories that can't be expressed this way. Might one of these answer every normative question? Consider the following line of reasoning:

Think back to Parfit's Miners Case. A complete normative theory would answer both the objectivist question, *what should Tanya do, objectively speaking?* and the subjectivist question, *what should Tanya do, given her uncertainty about where the miners are?* To do this, we could use a normative theory that's the conjunction of *two* theories of should, for example:

$$2N \quad \left\{ \begin{array}{l} S1 \text{ One should } O \phi \text{ iff } \phi \text{ing maximizes utility} \\ S2 \text{ One should } S \phi \text{ iff } \phi \text{ing maximizes expected utility} \end{array} \right.$$

Unfortunately, it's possible to imagine someone who's uncertain about S2. So this dual theory won't be able to answer questions of the form *what should so-and-so do, given her uncertainty about S2?* But that can be fixed! We can posit an infinite chain of more and more subjectivist theories. For example, we might have:

$$H \left\{ \begin{array}{l} S1 \text{ One should } O \phi \text{ iff } \phi \text{ing maximizes utility} \\ S2 \text{ One should } S1 \phi \text{ iff } \phi \text{ing maximizes expected utility} \\ S3 \text{ One should } S2 \phi \text{ iff } \phi \text{ing maximizes expected expected utility}^{44} \\ \text{Etc....} \end{array} \right.$$

A number of normative theorists have proposed something like this strategy.⁴⁵ And a number of others think that positing an infinite number of normative concepts *could* solve something like the Incompleteness Problem, but should not be done for other reasons, perhaps because positing an infinite number of varieties of normative concepts like SHOULD would be too radical a revision to make.⁴⁶

Let's call a normative theory that consists of an infinite series of recursively-defined theories of should a 'hierarchy theory.' A hierarchy theory could be designed so that the first (most objectivist) theory answers all questions of the form *what should one do, objectively speaking?* Then, the second theory could answer all the questions that the first couldn't answer: those of the form *what should one do, given one's uncertainty about the first theory?* The third would answer all the questions *this* theory can't answer, and so on. In such a theory, an infinite number of questions of the form *what should one do, given one's uncertainty about normative theory ____?* could be asked, but they'd all be answered by some normative theory or another.

That's the idea. The problem is that it's still possible to be uncertain about the whole hierarchy theory itself. For example, if our theory is the one described above, we might

⁴⁴ Assume that 'expected expected utility' bears the same relation to expected utility as expected utility bears to utility.

⁴⁵ The specific hierarchy I've given to is similar to one proposed in Spencer and Wells (Unpublished), section 8. Other examples include Fumerton's hierarchy of kinds of rationality in his 1989, page 120; Sepielli's 'orders of systemic rationality' in his 2013, pages 526-533; as well as the three kinds of hierarchy theories Smith discusses in her 2018, chapters 11-13: the multi-rule Hybrid system, the Expanded Moral Theory approach, and the Constrained Standards approach.

⁴⁶ See, for example, Lasonen-Aarnio 2015, Hedden 2015, and Kolodny and MacFarlane 2010.

imagine someone, call him Sid, who's uncertain between two definitions of utility. This could lead him to be uncertain between two versions of the theory. Neither version of the hierarchy theory could answer the question *what should Sid do given his uncertainty about which version is true?* because each would assume its own truth.

If our hierarchy theory can't answer a question, we can always supplement it with a theory that does. If we call our hierarchy theory 'H,' then a supplemented hierarchy theory, 'H+,' might be able to answer every question of the form *what should one do, given one's uncertainty about H?* But then there will be new normative questions that it can't answer: those of the form *what should one do, given one's uncertainty about H+?*⁴⁷

Uncertainty about a theory like *one should ϕ iff p* is either uncertainty about whether it's true that *one should ϕ iff p* (truth uncertainty), or uncertainty about whether p (application uncertainty). But what is uncertainty about a hierarchy theory like? Consider our sample hierarchy theory:

$$H \left\{ \begin{array}{l} S1 \text{ One should } \phi \text{ iff } \phi \text{ing maximizes utility} \\ S2 \text{ One should } S1 \text{ } \phi \text{ iff } \phi \text{ing maximizes expected utility} \\ S3 \text{ One should } S2 \text{ } \phi \text{ iff } \phi \text{ing maximizes expected expected utility} \\ \text{Etc....} \end{array} \right.$$

⁴⁷ The problem bears some structural similarity to the proof of Gödel's first incompleteness theorem (Gödel 1931). (Thanks to Larry Temkin for making this point.)

Gödel's first incompleteness theorem: Any consistent formal system F within which a certain amount of elementary arithmetic can be carried out is incomplete; i.e., there are statements of the language of F which can neither be proved nor disproved in F. (Raatikainen 2018)

The proof involves showing that it's possible to find a 'Gödel sentence' (a statement that can't be proved or disproved) for every possible system F. If we start with a system, F1, we will be able to find a Gödel sentence for it. And if we try to add this sentence as an axiom to make a new system, F2, we will be able to find a Gödel sentence for F2, and so on.

What do truth and application uncertainty about H look like? Application uncertainty is the least obvious, since there's no set of conditions such that, if it obtains, you've done what you should according to H. We could design a theory like this. It might say, for example, that you should act in accordance with the 'highest' theory in the hierarchy (the S_n with the lowest n , that is) that you have no uncertainty about. But such a theory could be expressed in the form *one should ϕ iff* _____. The hierarchy theories under discussion are not supposed to fit that form. They are pluralist theories in that they posit many different kinds of should, and it could be that you should do one thing according to one level and another according to a different level.

So, what does it mean to have uncertainty about hierarchy theories? Here's how I will characterize it:

Uncertainty about a hierarchy theory: uncertainty about every level in the hierarchy.

Truth uncertainty about a level: uncertainty about the *truth* of the biconditional.

Application uncertainty about a level: uncertainty about the *right side* of the biconditional.

As an example of truth uncertainty about H, consider Todd. Todd finds N compelling, but he's also pretty confident that pluralism about SHOULD is false. His confidence is split between H and some version of monism. Since he's not sure that 'shouldS1,' 'shouldS2,' and so on, are meaningful, he's uncertain about the truth of every level in H after the first.

As an example of someone who has application uncertainty about H, consider Tom, who's uncertain about expected utility. The expected utility of one of your options is a function of the probabilities (given your evidence) of each possible outcome (conditional on

your performing the option), and the utility of each outcome. So you might be uncertain about whether performing some option has the highest expected utility because: (i) you're unsure what your evidence is, (ii) you're uncertain what the correct probabilities to assign are, given your evidence, (iii) you're uncertain what the correct expected utility of each of your options is, perhaps because the calculation is complex, (iv) you're uncertain what all your options are, or (v) you're uncertain what the utilities of each possible outcome are.

Let's suppose that Tom is unsure what his evidence is. He thinks that his set of evidence is identical to the set of propositions that he knows. Unfortunately, he's not sure what exactly belongs in that set. Because whether or not ϕ maximizes expected utility or expected expected utility or... depends on what Tom's evidence is, Tom has application uncertainty about each level, beginning with S1.⁴⁸

Because it's possible to be uncertain about a whole hierarchy theory in a way that it can't be sensitive to, there are still subjectivist questions that hierarchy theories can't answer.

2.2.1.2. Other Kinds of Complexity

Not every normative theory with a more complex form is an infinite hierarchy of smaller theories. We might try a version of relativism instead. For example:

⁴⁸ What about Sid? In my example above, Sid is uncertain between two theories of utility, and so counts as being uncertain between two versions of the hierarchy theory (and about each individually). What kind of uncertainty does he have? It might seem obvious that he has application uncertainty, since he's uncertain about the right-hand side of the biconditional at every level. However, we can also describe him as being uncertain of the truth of the theory at each level, once we understand each instance of 'utility' as meaning something specific. His case shows that for hierarchy theories, some cases of truth uncertainty can be described equally well as cases of application uncertainty, and *vice versa*.

R *One should ϕ relative to some set of information iff ϕ ing maximizes expected utility, given that set of information.*

This looks promising because whenever someone is uncertain about the normative theory R, the state of uncertainty can be represented by some set of information. So the theory will deliver an answer to the question *what should one do, given one's uncertainty about R?* that's in some way sensitive to this uncertainty.

The trouble is that won't be sensitive in the right way, at least for truth uncertainty and certain kinds of application uncertainty. Why? It's primarily because the theory has more to it than relativity. R, for example, is an expected utility theory. And we can imagine agents who are uncertain about whether expected utility theories are true.

Suppose Tanya, in the Miners Case, is uncertain between R and a rival theory:

R* *One should ϕ relative to some set of information iff ϕ ing maximizes the chance of maximizing utility, given that set of information.*

Then there will be a set of information, I1, that encodes her uncertainty about where the miners are as well as her uncertainty about which of R and R* is true. It's clear what R and R* would say about what Tanya should do relative to an information set that only encodes her uncertainty about the location of the miners. R would say that she should block neither shaft, and R* would say that she should block one of the shafts (though it wouldn't care which). It's less clear what each theory would say about what Tanya should do relative to I1. Would the fact that she's uncertain about which theory is true make a difference or not? Luckily, we don't need to decide this to see that each theory will assume its own truth when answering. R will use I1 to calculate which option would maximize expected utility, and say that that is what Tanya should do; R* will use I1 to calculate which option would maximize

the chance of maximizing utility, and say that that is what Tanya should do. Neither will remain neutral about which of R and R^* is true.

What about application uncertainty? Recall Todd, who's uncertain about what belongs to his evidence set. He might want to know what he should do relative to the information set that's identical to his current evidence set. Let's suppose that this information set, I_2 , takes into consideration the fact that Todd is uncertain about what belongs to his evidence set. The trouble is that R will still rely on the truth about what exactly is in I_2 in giving its answer about what Todd should do relative to it.

We've considered normative theories with a more complex form than *one should ϕ iff* _____ in the hopes that their more complex structure could guarantee completeness. And it seems that they are no better equipped than simpler theories to deal with the problem posed by the Incompleteness Argument. But maybe there's more to the idea that we can define a normative theory so that it's guaranteed to be complete. Let's focus on this strategy directly.

My token trivial normative theory from section 2.1, *one should ϕ iff one should ϕ* , looks like a normative theory that has to be complete. For any question about what someone should do, it seems it will say the right thing. If you're concerned that it only says things about 'what one should do,' and not 'what one should do, objectively speaking,' or 'what one should do, given one's uncertainty about such-and-such,' we can change it to read:

The Expanded Trivial Theory:

one should ϕ iff one should ϕ

one should ϕ (objectively speaking) iff one should ϕ (objectively speaking)

one should ϕ (given one's uncertainty about such-and-such) iff one should ϕ (given one's uncertainty about such-and-such)

Etc.

Written this way, it must say the correct things in response to all questions of the form *what should one do, given one's uncertainty about such-and-such?* This includes questions of the form *what should one do, given one's uncertainty about the Expanded Trivial Theory?*

The obvious difficulty of using theories like this to respond to the Incompleteness Argument is that the Expanded Trivial Theory hardly seems like a normative 'theory' at all. It is completely uninformative and explains nothing. If the only complete normative theories are those that are uninformative and unexplanatory, then that's just as concerning as the conclusion that no normative theories could be complete.

Another possibility is a normative theory that's somehow an 'infinitely-supplemented' hierarchy theory. Such a theory might begin with an infinite hierarchy of theories of should, and then contain another theory of should that's sensitive to uncertainty about the hierarchy, and another that's sensitive to uncertainty about the new theory (the hierarchy plus the extra theory), and so on. Such an all-encompassing theory would also have to be designed so that if you were ever to try to refer to, or be uncertain about, the whole theory, you'd somehow always just fall short. A theory like this would seem to be inexpressible.⁴⁹ But could it be complete? I'm not so sure. I'm troubled by Todd's kind of uncertainty: uncertainty between pluralist hierarchy theories and monist normative theories. If you have this kind of uncertainty, then it seems that you would (or could) be uncertain about every level in the hierarchy, simply because that level belongs to a hierarchy.

⁴⁹ Thanks to David Black for suggesting the possibility of a normative theory that's complete but inexpressible.

The lesson here seems to be that the kind of theory that we'd want to be able (in principle) to build or find could not be complete. Trivial ones might be complete, but they'd be uninformative and so not useful to us. We might also be able to gesture vaguely toward a possible substantive normative theory that's complete. But this is little better than pointing out that the set of normative facts is complete, even if we could never build or find a theory to cover them all.

In my explanation of the Incompleteness Argument in Section 1.1, I assumed that the only normative theories under consideration were theories about what agents *should* do. But there are many other normative concepts that we'd expect a complete normative theory to answer questions about. For example, there's what it's *rational* for agents to do, what it would be *best* for them to do, what they have *reason* to do, what they are *obligated* to do, and so on. Might an all-encompassing normative theory that includes accounts of all these notions—and not just *should*—have a better shot at completeness? If every theory of *should* leaves some questions unanswered, might theories of, say, rationality or blamelessness take up the slack?⁵⁰

Unfortunately, since we've already seen that helping ourselves to an infinite number of varieties of *should* doesn't work, it's not clear why this approach would be any better. Note that we could simply call one of our subjective *should* notions 'rationality' to get the following two-level theory:

⁵⁰ There are also other normative concepts that don't have to do with agents and their options. For example, there's what's *valuable*, what's *good*, etc. But it's less obvious how appealing to these concepts could help.

$$N \left\{ \begin{array}{l} S1 \text{ One should } \phi \text{ iff } \phi \text{ing maximizes utility} \\ S2 \text{ It's } \textbf{rational} \text{ for one to } \phi \text{ iff } \phi \text{ing maximizes expected utility} \end{array} \right.$$

If this is the route we take, we can count S2 as being able to answer at least some questions of the form *what should one do given one's uncertainty about S1?* Obviously it won't be able to give answers of the form *one **should** do such-and-such given one's uncertainty about S1*, but it will be able to give answers that are appropriately responsive to what's being asked. If we imagine for a moment that the theory S2 is really a person who believes that S2 is true, she might say: "Oh, I think you mean to say '*what's it rational to do given one's uncertainty about S1?*' and the answer is: *it's rational to ____ given one's uncertainty about S1.*"⁵¹

So appealing to a notion like rationality or blamelessness could work just as well as appealing to a more subjectivist notion of should. But to escape the Incompleteness Argument, it would have to work better. Whatever normative concepts we appeal to in order to answer the subjectivist questions that other theories can't answer, there will still be more subjectivist questions without answers. For example, we can ask *what's it rational for one to do, given one's uncertainty about S2?* We can always appeal to a third concept to answer this question—for example, 'blamelessness'—but this strategy is no more promising than a hierarchy theory composed entirely of theories of should.

⁵¹ If there are questioners who resist ("No, I meant to ask *what **should** one do given one's uncertainty about S1?*"), they do not concern us here. The question I'm expressing by the phrase *what should one do given one's uncertainty about S1?* is not meant to be theory-laden or otherwise especially attached to the specific normative notion *should*.

2.2.2. 'Fastidious' Responses

2.2.2.1. Answers

One objection to Premise 2 of the Incompleteness Argument is that it's actually easy for a normative theory to be complete. It's easy because normative theories *can* answer questions like *what should one do given one's uncertainty about p?* without being sensitive to one's uncertainty about p. On a popular account of questions and answers, a question places a partition over the space of possible worlds, and an answer to it rules out all but one cell in the partition.⁵² For example, my question *do you own Bitcoin?* divides the space of possible worlds into ones where you own Bitcoin and ones where you don't. And your answer *yes, I do* rules out the cell in the partition containing the worlds in which you don't own Bitcoin.⁵³

If this is the right account of questions and answers, then the theories that I've claimed can't answer certain questions actually can. For example, the subjectivist question *what should Tanya do given her uncertainty about where the miners are?* may divide the space of possible worlds into ones where she should block shaft A, ones where she should block shaft B, and ones where she should block neither shaft. According to an objectivist theory like *one should ϕ iff ϕ ing maximizes utility*, Tanya should block shaft A. (Let's assume that's where the miners are.) Since this response rules out all but one of the cells in the partition, the objectivist theory does count as giving an answer to the question.⁵⁴

⁵² See, e.g., Groenendijk and Stokhof 1984.

⁵³ Likewise, my question *where's Henry?* divides the space of possible worlds into cells that represent places Henry could be. A response like *Henry's in town* counts as a partial answer if it rules out some of these cells, but not others. If I want to know where Henry is in town, exactly, I can ask a more specific question.

⁵⁴ Questions that involve counterfactuals might seem to work like questions with the form I've used to express subjectivist questions (*what should one do, given one's uncertainty about such-and-such?*). Consider the following question and answer pair:

Q. If we were in Australia right now, would it be dark out?

The strength of this objection rests on a particular account of answers. But the account is plausible. And even if it's not exactly right, other accounts of answers could have the same upshot. So I happily concede that the objection can work against the Incompleteness Argument as it's presented in section 1.1.

The trouble is that the notion of 'answering a question' can be replaced by something else to create the same problem. Here's another way of stating the Incompleteness Argument:

The Incompleteness Argument (2)

1. A complete normative theory would be able to **adequately answer** every normative question.
2. For any normative theory, N, there's a question of the form *what should so-and-so do given her uncertainty about N?* that N can't **adequately answer**.
3. Therefore, no normative theory could be complete.

A. We're not. It's light out here. (And that's all that matters.)

This answer is uncooperative. The person giving it refuses to consider what would be the case if she and the questioner were in a different place. So it may be possible to claim that this response does not count as an answer at all. If the question puts a partition over only that portion of possibility space in which both speakers are in Australia, then the response would not rule out any cells in the partition.

Because subjectivist questions also involve a qualification portion that can be ignored by the objectivist, it might seem that there's a way to prove that, even on the partition view, objectivist theories can't answer subjectivist questions. Unfortunately, this doesn't work. The difference is that while 'If we were in Australia right now' might restrict the space of possible worlds, 'given one's uncertainty about p' doesn't do this. The subjectivist is not looking for an answer in which all the worlds in which one isn't uncertain about p are ruled out; she's looking for an answer that's *given in the right kind of way*.

This version of the argument requires an account of ‘adequate answer.’ And it must be one such that: no adequate answer to a question of the form *what should one do given one’s uncertainty about p?* could be given by a theory that relies on the truth about whether p. Here’s a possible account:⁵⁵

An **adequate answer** to a question is an answer that’s appropriately sensitive to the intention behind the question that’s asked.

When one asks *what’s it rational for Tanya to do, given her uncertainty about where the miners are?* one expects an answer that isn’t determined by the actual location of the miners. Why? There seems to be an important theoretical role to be filled by theories that don’t depend on what agents are uncertain about in saying what they should do.

A version of the Incompleteness Argument using ‘adequate answer’ could work just like the original. No normative theory would be able to give an adequate answer to questions about what one should do, given one’s uncertainty about that theory itself.

⁵⁵ What about a question like *what should one do, given one’s number of freckles?* Such a question might falsely presuppose that the number of freckles you have makes a great difference to what you should do. The questioner’s intention is to receive an answer that is ‘appropriately sensitive’ to numbers of freckles. (Thanks to Andy Egan for this point.) I’m tempted to say one of two things here. (1) An adequate answer to this question would be sensitive to numbers of freckles, but a complete normative theory wouldn’t have to answer this kind of question because there are some limits on the types of normative questions that it must answer, or (2) an adequate answer to this question would not have to be sensitive to numbers of freckles, and could instead be something like: “one’s number of freckles is (almost always) irrelevant to what one should do.” The main point is that if (as Parfit’s Miners Case and others seem to show) there are subjective normative facts that can be expressed by phrases like ‘one should do *blah* given one’s uncertainty about such-and-such,’ then a complete normative theory would be expected to accommodate them (or perhaps ‘explain’ or ‘deliver’ them). And a theory can’t do this if it ‘answers’ questions of the form *what should one do, given one’s uncertainty about whether p?* in a way that relies on the truth of whether p, even if we decide to treat these responses as answers.

2.2.2.2. Questions

Premise 1 of the Incompleteness Argument says that a complete normative theory would (adequately) answer all normative questions. But would it really have to answer all normative questions? There might be some normative questions that it makes no sense to require answers to. Consider: *what should Alice do given her uncertainty about N?* asked when Alice is not uncertain about N at all. Or *what is the best possible world?* when there is no best possible world.

We *could* still require a complete normative theory to answer these questions, though the answers would be things like: *Alice is not uncertain about N (though if she were, she should do x), there is no best possible world*, etc. But there might still be some questions that we wouldn't expect any complete normative theory to answer. For example: *If mercury weren't poisonous, should I eat it?* Counterfactual questions that don't provide enough information to know what the world would be like may not be possible to answer. If mercury weren't poisonous, would it be nutrient? Would it have any health benefits or risks? Without this information, it's unclear how anyone could answer. Or consider a question like: *What should I do according to normative theory N47?* (asked when theory N47 is not the correct normative theory and would give conflicting advice from the correct one). While I may want to know the answer to this question, I'm not sure that we should require a complete normative theory to answer it.

However, we can again turn to a more qualified version of the Incompleteness Argument:

The Incompleteness Argument (3)

1. A complete normative theory would be able to **adequately answer** every **reasonable** normative question.
2. For any normative theory, N, there's a **reasonable** question of the form *what should so-and-so do given her uncertainty about N?* that N can't **adequately answer**.
3. Therefore, no normative theory could be complete.

Even if there are some normative questions that it makes no sense to require any complete normative theory to answer, questions of the form *what should one do, given one's uncertainty about such-and-such?* still do require answers. Unlike the question about mercury, we often have intuitions about what would count as better and worse answers to them. And unlike the question about theory N47, it seems reasonable to require answers to them from any normative theory that claims to be complete.

2.2.3. 'Dismissive' Responses

I've been arguing that the Incompleteness Argument (or some version of it) succeeds. But does it really matter if no normative theory could be complete? Consider a final objection:

The Objectivist Response

What the Incompleteness Argument shows is that no normative theory could possibly accommodate all subjectivist intuitions. But that's only a problem for the subjectivist.

It's perfectly fine to end our theorizing with a theory that doesn't answer all subjectivist questions (or that doesn't answer them all 'adequately').

We might read 'it's perfectly fine to end our theorizing' in a practical sense. If nothing else, the universe will end at some point and we'll have to stop then. And it would be fine for us to do so. But the problem raised by the Incompleteness Argument is not a practical one. If it's right, there could never, even in principle, be a point at which we have a normative theory that does everything that one would be expected to do: answer all normative questions.

The objectivist must explain why it's fine for normative theories to do less. And so far, debates between objectivists and subjectivists have concerned narrower kinds of normative theories, where the objectivist position is easier to defend. As I note in section 2.2.1.2, if we're only concerned with a theory of the normative concept *should*, then we might be fine with an objectivist theory. Subjectivist questions could always be left to be covered by other normative notions like 'rationality,' 'blamelessness,' 'criticisability,' etc. But narrow objectivist normative theories like this owe their plausibility to the fact that they're not meant to be the *only* normative theories.⁵⁶

Here's a more direct response to the objection. First, it does not seem okay to end our theorizing with a normative theory that only answers objectivist questions. In particular, we

⁵⁶ As evidence of this, normative theorists who take themselves to be objectivists and to defend objectivist theories of rationality, should, etc., often appeal to other normative concepts when confronted with counterintuitive implications of their theories. For example, Lasonen-Aarnio 2014 defends an objectivist (or 'externalist') account of epistemic rationality according to which, in a case of higher-order defeat, what's rational for you to believe may be something that you have good reason to think is irrational. Lasonen-Aarnio recognizes that there's still something "unreasonable" or "criticizable" about continuing to hold the belief, but makes the case that we still need an objectivist account of *rationality*.

seem to require a theory that can answer some questions like *what should one do given one's uncertainty about the objectivist normative theory?* Consider what would happen if we tried to stop with a theory like *one should do whatever maximizes utility*. We'd still have subjectivist questions, say, about what Tanya and Susan 'should do given their situations,' what they'd 'be blameless for doing,' or however we choose to phrase them. And the fact that we can't develop a theory to answer all subjectivist questions isn't a good reason to settle with a theory that answers none of them.

Second, if we require a normative theory that can answer some questions of the form *what should one do given one's uncertainty about normative theory N?* then we should want a theory that can answer all of them. There's no principled stopping point. Even a subjectivist might resist the idea that a final normative theory would have to answer *every* possible subjectivist question of the form *what should one do given one's uncertainty about normative theory N?* But the trouble is that there doesn't seem to be a good theoretical reason to stop trying to answer these questions. Whatever reason we have to begin a hierarchy theory by adding a second level is also a reason to add another and another. There may be a practical reason to stop if we, say, run out of resources. But the theoretical reason to keep going is that there are always more normative questions that call for (adequate) answers.

Here ends my case for the Incompleteness Argument. I think something close enough to it to be concerning must be right. If no normative theory could be complete, then no normative theory could make sense of all of the subjectivist normative facts about what one is to do given one's uncertainty about something. And any theory that doesn't do this will fall short: we'd still want more normative theory in order to accommodate them.

3

Starting an Open-Ended Normative Theory

In this final chapter, I take for granted the conclusions of chapters 1 and 2. I ask: *What are we to do if a complete normative theory is unattainable for the reasons given in Chapter 2?* And in answering, I help myself to the Normativity *Lite* Framework and the all-things-considered SHOULD concept proposed in Chapter 1.

If it's impossible to ever build a complete normative theory, what should we aim to do instead? I'll argue that our best option is to build, and continue to build, a normative theory that can be extended indefinitely. In section 3.1, I explain what such an 'open-ended' theory would be like and why I think it's our best strategy. But most of this chapter is devoted to proposing a good 'first stage' in an open-ended theory. If there is no such thing as a best normative theory, then finding a good first stage in an open-ended one is an important provisionally 'end' goal for normative theorists to have now. In section 3.2 I describe some theoretical virtues of good first stages. In section 3.3 I explain why a hierarchy theory could make a good first stage, and in section 3.4 I propose a specific one to use: the '*Lite* Hierarchy Theory.' In section 3.5 I apply the *Lite* Hierarchy Theory to the problems of moral uncertainty and higher-order defeat to show how it might help to solve them, and also to show some of its limitations. In section 3.6 I discuss how the next stages could be designed.

3.1. The Idea of an ‘Open-Ended’ Normative Theory

Imagine the set of normative facts as a two-dimensional space, which we can call ‘normative space.’⁵⁷ And imagine normative theories as ‘covering’ portions of this space. For example, the following diagram represents the fact that a theory of *Hilda’s* reasons (e.g. *R is a reason for Hilda to ϕ just in case R counts in favor of Hilda ϕ ing*) is neither a complete theory of reasons nor a complete normative theory. It does not completely ‘cover’ either of these regions of normative space:

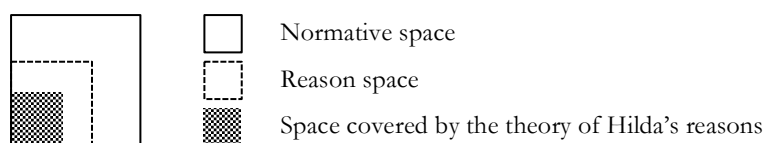


Figure 2. Normative space covered by an incomplete theory

A theory ‘covers’ a region of normative space if it accommodates, or explains, or delivers all normative facts in that region. More precisely, I’ll say that a theory covers a region of normative space if it makes (consistent) general claims that could be used (in combination with other, ideally non-normative, facts) to derive propositions expressing each normative fact in this region, or else the negation of these propositions.⁵⁸ A ‘normative theorist,’ then,

⁵⁷ The set of normative facts is so huge and complex that it could probably be represented in many more dimensions than two. I use two because it’s all that’s needed here.

⁵⁸ And to derive them in the appropriate way, if there is one. ‘Subjective normative facts’ (i.e. answers to subjectivist normative questions) do not count as being covered by normative theories that merely ‘get them right,’ but aren’t appropriately sensitive to the agent’s epistemic state. In other words, if a normative theory cannot (adequately) answer a subjectivist normative question, then it does not cover a corresponding portion of normative space.

is one who finds or develops theories (e.g. general principles) that cover portions of normative space.⁵⁹

Normative theorists engage in many different projects. Some develop ethical theories, theories of rational belief, theories of good decision-making, etc. The aim of each project is to find a theory that covers some portion of normative space. It's plausible, then, to see the aim of normative theorizing in general as that of finding an 'ultimate' normative theory that accurately covers all of normative space. We can represent such a theory like this:

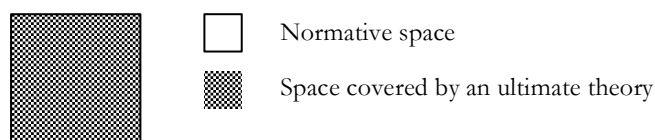


Figure 3. Normative space covered by an ultimate theory

In Chapter 2, I argued that there could be no complete normative theory. The argument went as follows: a complete normative theory would (adequately) answer every normative question, it's possible to be uncertain about every normative theory, no normative theory could (adequately) answer normative questions of the form *what should one do, given one's uncertainty about [the normative theory]?* and so no normative theory could be complete.

If no normative theory could be complete, then there can be no ultimate normative theory. If no normative theory could (adequately) answer every normative question, then no normative theory could cover all of normative space, let alone accurately cover all of

⁵⁹ Normative theories must posit some normative principles or otherwise make some interesting generalizations. This rules out versions normative particularism as normative theories. (Normative particularism is the view that there can be no general normative principles that apply to every case; even to every case of some kind. See, e.g., Dancy 2004.) Here, particularists count as skeptics about the existence of true normative theories.

normative space. For every question that a normative theory can't answer, there is a normative fact (the answer) that it doesn't cover.

If there can be no ultimate normative theory, then what should the aim of normative theorizing be? If we can find a theory that accurately covers region A of normative space, and another that accurately covers region A (in the same way) as well as region B, then the second theory is better. The problem posed by the Incompleteness Argument is that there's no such thing as a *best* normative theory. Instead of an ultimate theory, there could be an infinite series of better and better incomplete theories that (accurately) cover more and more normative space.

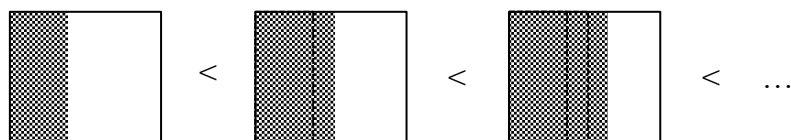


Figure 4. Normative theories that cover successively more space⁶⁰

The problem could be worse. Consider Bartha, Barker, and Hájek's puzzle of Saint Peter's Offer.

You are met by Saint Peter as you arrive at the gates of Heaven. Unfortunately, Saint Peter tells you, it has not been settled how long you can stay. In acknowledgment of a few acts of kindness that you once performed, he invites you to write down a positive integer as large as you please, stay in Heaven for that many days, and then return to Hell—as your initial period of thinking and writing is also to take place in Hell. To keep things simple, Saint Peter asks you to use tally notation (1 stroke = 1 day in Heaven).⁶¹

⁶⁰ Figure 4 is a bit misleading. It also shows theories that cover greater and greater *percentages* of normative space. But if normative space is infinitely large, this need not be the case.

⁶¹ Bartha, Barker, and Hájek 2014 page 631.

Since there is no largest number of strokes you can make (you're in Hell, so you won't die), it seems that you are doomed to make an irrational choice. Whatever number of strokes you choose, you could easily have made just one more, and that would have been a better thing to do. But the especially puzzling thing about this case is that if you continue to make strokes indefinitely, you will stay in Hell forever—the worst possible outcome.

The problem for normative theorists is not as bad as this. Normative theorizing is not done in Hell, and if it were possible to continue to add to a normative theory forever, the theory could just continue to get better.

But the fact that we would never reach a best normative theory should make us wonder what the aim of normative theorizing could be. The end goal, it might be said, is either a complete theory, or an incomplete theory. Aiming for an incomplete theory is like choosing to stop after some specific number of strokes. This can't be the end goal of normative theorizing. An incomplete normative theory will leave more normative space to be covered and hence more theorizing to be done. However, aiming for a complete theory is like aiming for a largest number. Since there is no such thing, the project of normative theorizing could look misguided.

But even if there's no such thing as a complete theory, normative theorists do not have to aim, ultimately, for any specific incomplete theory. If there's no best normative theory to build, we can ask instead: *what's the best **kind** of theory a normative theorist could build?* And the right answer, I think, is: a normative theory that can be extended indefinitely. This is the best *kind* of normative theory since it wouldn't put an arbitrary stop to normative theorizing.

An 'open-ended' normative theory can be represented like this:

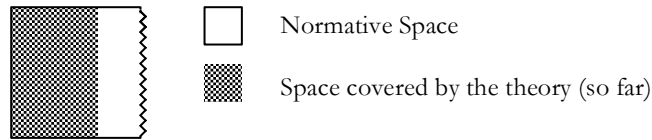


Figure 5. *An open-ended normative theory*

The jagged line represents the fact that the theory could never grow to cover all of normative space. This is what makes the theory open-ended as opposed to simply incomplete. An ‘open-ended’ theory is one that (a) is meant to be extended, and (b) will always be possible to extend further. The difference between open-ended and ‘closed’ normative theories is meta-theoretical. A closed complete theory claims to cover all of normative space, and assumes that this is possible. A closed incomplete theory claims to cover some smaller portion of normative space. Neither theory claims that it can be extended to cover more space. An open-ended theory claims that it is meant to be extended, and also that it can be extended indefinitely: that it’s impossible to completely cover normative space.

Open-ended normative theories are built in *stages*. So a visual representation of one never shows ‘the whole thing’—it just shows some number of *stages* of one. The following might represent the same open-ended normative theory with an extra stage:

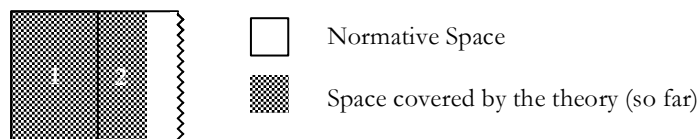


Figure 6. *An open-ended normative theory with two stages*

A stage in an open-ended theory is just a smaller theory. And any two consistent theories can be joined to create a larger one. But it's often convenient to distinguish between smaller theories in an open-ended one.

Consider some 'two-level' normative theory: a conjunction of an objectivist normative theory and a subjectivist normative theory. The theory posits a subjective should concept, according to which, what an agent subjectively should do is determined by facts that are sensitive to the agent's epistemic state. If I have no idea which numbers will win the lottery, then the (objective) fact that 11, 28, 31, 46, 59 is the winning combination cannot determine whether I subjectively should choose these numbers. Perhaps, if I'm to play the lottery, I subjectively should choose numbers at random. The theory also posits an objective should concept, according to which, what an agent objectively should do can be determined by any and all kinds of facts. For example, the fact that the string of numbers above is a winning combination may determine that I objectively should choose them, if I'm to play.

This theory can be represented in two ways. It would normally make sense to represent it as a closed theory:

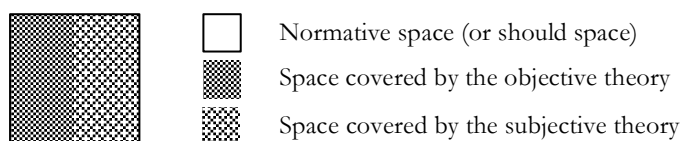


Figure 7. A closed two-level normative theory

But we can also represent it as an open-ended one:

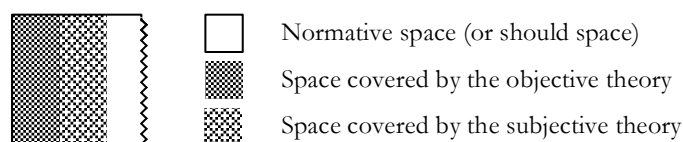


Figure 8. *An open-ended two-level normative theory*

In the first diagram, it didn't matter how I arranged the 'coverage areas' of the two theories, so long as they covered the whole of normative space. However, in this second diagram it does matter that I've put the objectivist theory on the left, or 'first,' and the subjectivist theory directly to the right of it, or 'second.' The jagged line represents the direction of 'growth' of the open-ended theory, and the uncovered space represents normative facts that are even more subjective than those in the light gray area. These include normative facts about what one should do given one's uncertainty about the subjective theory, or about the two-level theory as a whole. We could also represent a two-level theory as a single stage in an open-ended normative theory and represent each of its levels as sub-stages.

3.2. A Good First Stage in an Open-Ended Normative Theory

In this section, I'll describe some theoretical virtues of *first stages* in open-ended normative theories. While there's strictly speaking no such thing as a 'best' first stage, there are definitely respects in which they can be better and worse.

One good thing for a first stage in an open-ended normative theory to do would be to be as complete as possible in the respects or 'dimensions' in which it can. For example, it would look like this:

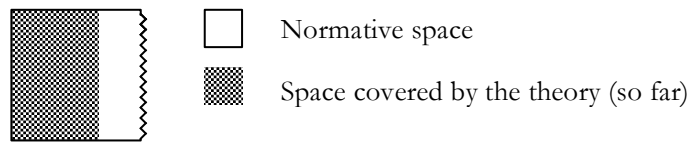


Figure 9. A good first stage in an open-ended normative theory

as opposed to this:

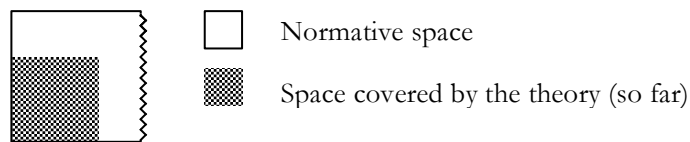


Figure 10. A worse first stage in an open-ended normative theory

or this:

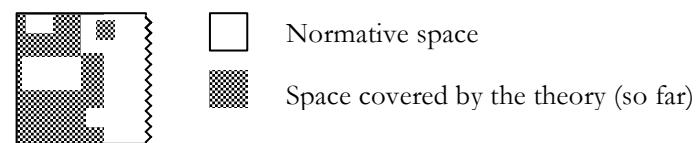


Figure 11. An even worse first stage in an open-ended normative theory

In other words, one virtue of good first stage in an open-ended theory is that of:

1. **Being organized and complete (where completeness is possible).** Only leaving uncovered normative space on the right.

If a first stage in an open-ended theory leaves a lot of normative space uncovered that could be covered by a different theory, then it isn't an example of one that's best in all respects in which it's possible to be best.

A second virtue is that of:

2. **Going far enough.** Covering a sufficient amount of normative space in the rightward direction so that most of the normative questions we currently care about are answered.

A good first stage in an open-ended theory would be at least as complete as other normative theories that we already have.

A third virtue is that of:

3. **Being unified.** Not being a patchwork of unrelated pieces of theory that cover different portions of normative space.

Unification is a general theoretical virtue. We'd expect a good first stage to be unified to the extent that unification is possible. A fourth virtue is that of:

4. **Accuracy.** Not giving the wrong answers to any normative questions.

This one is obvious. A theory that gets the wrong results won't accurately cover that part of normative space. And there's no reason that an open-ended theory would have to get any normative questions wrong, so a good first stage in one won't either.

And partly because of this, a good *description* of a good first stage would also:

5. **Avoid being contentious.** Avoid taking sides on controversial issues.

The first four theoretical virtues are the only ones really applicable to *first stages* of open-ended theories. The first two are structural: they're about the amount of coverage that a first stage has. Continuing with the visual metaphor, the first virtue is about coverage up, down, and to the left, while the second is about coverage to the right. The third and fourth are about the content of the theory: its accuracy and how its component parts fit together.

Virtues 5 and 6 don't apply to first stages themselves so much as to good proposals, sketches, or descriptions of first stages. There might be some true open-ended theory that accurately (and adequately) answers every normative question in its purview, but is extremely contentious. Perhaps, as a matter of fact, all forms of consequentialism are false. Then this would be reflected in the first stage of any true open-ended theory. However, my goal here is more abstract. It's that of explaining how we could start an open-ended normative theory. So I'm hoping to find a neutral way of describing one. Since many first-order normative questions are still undecided, and since nothing about the idea of an open-ended theory presupposes specific answers to them, there's no need to saddle an explanation of *how to make a good start to one* with more commitments than necessary.

However, theorists face a tradeoff between neutrality and informativeness. So, a good proposal for or sketch of a good first stage would also:

6. **Be precise and informative enough.** Say enough to be interesting and helpful for those wanting to fill in the details.

For example, a theory (or ‘theory’) like: *one should always do whatever the correct theory says one should do* is extremely neutral. But it doesn’t say anything helpful. It doesn’t explain anything or tell us anything we don’t already know.

3.3. Infinite Hierarchy Theories as Good First Stages

Infinite hierarchy theories begin like two-level theories, but posit an infinite number of levels as well as a rule for generating each new level from the one before. Because of this, they promise to cover infinitely more normative space ‘in the rightward direction’ than two-level theories. Since they can cover so much space in such a principled manner, they seem ideally suited to have virtues 1, 2, and 3.

As an example of a simple infinite hierarchy, consider Richard Fumerton’s. Fumerton describes the following puzzle for accounts of rational action:

Fumerton’s Puzzle

On the one hand, we feel the force of allowing that an action is rational provided that the agent justifiably believes that the action is rational. On the other hand, on *any* analysis of rational action it looks as if an agent can have a justified but *false* belief that his action is rational.⁶²

To understand this puzzle, first consider someone who acts as she believes is rational. And imagine that she’s totally justified in believing that her actions are rational. For example, suppose that Angie believes that it’s rational for her to shut down her computer every night and restart it every morning. She believes this because she’s been told by reliable sources that doing so will save electricity and extend the life of her computer. Even if these things

⁶² Fumerton 1990 page 119.

aren't true, Angie seems to act *rationally* when she shuts down her computer every night. And the same seems to hold, more generally, for anyone else who acts in a way that he or she justifiedly believes is rational.⁶³

But this causes a problem if it's always possible to be justifiedly mistaken about what's rational. Then, for any theory of rationality you can devise, it will be possible to find a case where: it's rational for an agent to do one thing according to the theory, but it's actually rational for her to do something else given her false justified belief about what's rational. Fumerton seems to have found a method for generating counterexamples to any proposed theory of rationality.

Fumerton considers and rejects a number of ways of solving the puzzle.⁶⁴ His solution is to conclude that "there are fundamental and important ambiguities in the concept of rational action."⁶⁵ And he suggests that for any analysis of rationality, R, there will be an infinite number of 'derivative' senses of rationality. In addition to the sense R, according to which an action is rational just in case it meets the conditions for being R, there will be: (i) another sense, R', according to which an action is rational just in case the agent has a justified belief that the action meets the conditions for being R, (ii) a sense, R'', according to which an action is rational just in case the agent has a justified belief that the action meets the conditions for being R', ... and so on.⁶⁶

Fumerton's Infinite Hierarchy

Level 1 It's rational for A to ϕ just in case conditions C are met.

⁶³ See, e.g., Way and Whiting 2016.

⁶⁴ Fumerton 1990 pages 119-120.

⁶⁵ *Ibid.* page 120.

⁶⁶ *Ibid.* page 121.

Level 2	It's rational' for A to ϕ just in case A has a justified belief that conditions C are met.
Level 3	It's rational'' for A to ϕ just in case A has a justified belief that A has a justified belief that conditions C are met.
...	...

Introducing an indefinite number of different kinds of rationality like these solves the puzzle. If an agent has a justified but false belief that her action is rational in some specific sense, then it obviously won't be rational in that sense, but it will be rational in some *other* sense. Fumerton explains that "there is no need to choose between these different derivative concepts of rational action" and that his framework "allows us to generate derivative concepts as needed."⁶⁷

To see his theory in action, suppose we choose a specific theory of rationality: *it's rational for you to do something just in case it will best satisfy your desires*. Then, the hierarchy we get is:

Fumerton's Infinite Hierarchy + DS

Level 1	It's rational for A to ϕ just in case ϕ ing would best satisfy A's desires.
Level 2	It's rational' for A to ϕ just in case A has a justified belief that ϕ ing would best satisfy A's desires.
Level 3	It's rational'' for A to ϕ just in case A has a justified belief that A has a justified belief that ϕ ing would best satisfy A's desires.
...	...

⁶⁷ *Ibid.* page 124.

In many cases, this hierarchy theory says plausible things about what it's rational to do. For example:

Theodore can choose between a skateboard or a bike for his birthday. He chooses the skateboard because he justifiedly believes that it will make him happier. But then he has a painful accident on the skateboard and regrets his decision. And he's right to regret it: he would have been much happier if he'd chosen the bike.

According to Fumerton's Hierarchy + DS, it was not rational for Theodore to choose the bike. But it was rational' for him to choose the bike. So far so good. Now consider a different case:

Marcie is waiting for an important package to arrive. It's her renewed passport. She must sign for it, and she needs it for a trip she plans to take the next morning. However, she also knows that today is the last day the museum in town is showing a free exhibit that she wants to see. She knows that her desires would be best satisfied by going to the exhibit at just the right time to be home to collect her passport. And she knows that one of two actions would guarantee this: she could (a) go to the museum in the morning, or she could (b) go to the museum in the afternoon. If the package arrives in the morning, then (b) would best satisfy her desires. If the package arrives in the afternoon, then (a) would best satisfy her desires. She doesn't believe that missing the exhibit entirely and staying home all day to wait for the delivery is the option that best satisfies her desires. And yet, that's exactly what she does in order not to risk missing the package (which in fact arrives at 4pm after the museum has closed).

According to Fumerton's Hierarchy + DS, it's not rational for Marcie to stay home all day instead of visiting the museum. (She could have gone to the museum in the morning and still been home to get her package in the afternoon.) But it's also not rational' for her to do this. Nor is it rational in any other sense, since she hasn't formed any of the requisite higher-order beliefs. (And if she had considered the relevant propositions, she wouldn't have believed them either. She'd just have a justified belief that ... she has a justified belief that staying home all day is not the action that would best satisfy her desires.) But this can't be the right result. If Theodore's decision is rational in at least *some* important sense, then so is Marcie's.

Cases like Marcie's show that Fumerton's Hierarchy is wrong to require an agent to justifiedly believe that an action meets the conditions of the previous level in order for that action to be rational in some sense. So it doesn't have the fourth virtue of good first stages in open-ended normative theories: that of not saying things that are false.

But other hierarchy theories have been proposed. Andrew Sepielli, for example, in his "What to Do When You Don't Know What to Do When You Don't Know What to Do..."⁶⁸ suggests that there could be an indefinite number of 'orders' of rationality. And he gives the following description of what these orders are like:

What it's N+1st Order Rational to do depends upon the agent's beliefs or credence whose contents are propositions about a kind of value V, where V depends upon the way the world is according to expressions of the beliefs or credences upon which Nth Order Rationality depends (unless N = 1, in which case it depends on the agent's beliefs or credences regarding objective normativity).⁶⁹

⁶⁸ Sepielli 2014.

⁶⁹ *Ibid.* page 530. This is Sepielli's 'C3' formulation.

This description is very general, and doesn't tell us enough to derive necessary and sufficient conditions for each order of rationality. Because of this, it's lacking with respect to the sixth virtue: informativeness. But Sepielli's description of a hierarchy theory makes a better (description of a) first stage in an open-ended theory than Fumerton's because it avoids saying things that are false or contentious.

Fumerton's Hierarchy isn't even a possible way of filling in the details in Sepielli's account, since it doesn't involve beliefs about any kind of value. But there are other ways of fleshing out Sepielli's account. Jack Spencer and Ian Wells have proposed a hierarchy theory of rationality that's motivated by cases involving agents who are uncertain about even more than Marcie is. The two contend that it's possible to be uncertain about what the expected values of your options are, and that in these cases it can be rational for you to choose an option that will maximize neither value nor expected value, but *expected expected value*. Here's their case:

Fire: The fire alarm rings, and the agent, a firefighter, hurries into the truck. On the ride over she deliberates. She knows that she will have three options: she can enter the building via door A, door B, or door C. Since she does not know the exact distribution of residents in the building, she does not know which option will result in the most rescues. Based on her credences about the distribution of residents, she calculates the expected values of each of her options, and writes the values on a card. After exiting the truck and attaching the hose, she races toward the building and reaches into her pocket. But the card is gone. And time is of the essence: she knows all of the residents will die in the time it would take her to recalculate the expected values. She remembers that the expected value of entering via door B is 9. Of the other two doors, she remembers that one has an expected value of 0, and that the other has an expected value of 10, but she cannot remember which expected value goes with which door. (In fact, entering via door C has an expected value of 10, as the lost card attests.)⁷⁰

Spencer and Wells claim that in this case:

⁷⁰ Spencer and Wells (Unpublished) page 12. A slightly more complex version of the case also appears in their 2017 article "Why Take Both Boxes?" on pages 7-8.

An agent facing Fire, though unable to access either actual value or expected value, is able to access *expected* expected value where, just as expected value is a credence-weighted average of the agent's hypotheses about actual value, expected expected value is a credence-weighted average of the agent's hypotheses about expected value. An agent facing Fire is rationally required to enter the building via door B, and she is, we believe, because she is rationally required to maximize expected expected value.⁷¹

Here is a version of the kind of hierarchy theory they propose (simplified a bit for present purposes):

The S&W Infinite Hierarchy

Level 1	It's rational for A to ϕ just in case ϕ ing would maximize value.
Level 2	It's rational' for A to ϕ just in case ϕ ing would maximize expected value.
Level 3	It's rational'' for A to ϕ just in case ϕ ing would maximize expected expected value.
...	...

This theory is able to get the right result about Marcie. While staying home all day doesn't count as rational, it does count as rational'. That's because the expected value of leaving the house during the day to go to the museum is low compared to the expected value of staying home. And it can handle more complex cases of uncertainty, like the case Fire.

It's also a more specific version of Sepielli's account. In it, each higher level (or 'order') depends on the agent's credences about a specific kind of value: expected (or expected expected...) value. And this value does 'depend on the way the world is according to

⁷¹ *Ibid.*

expressions of the beliefs or credences upon which the next-higher kind of rationality in the S&W Infinite Hierarchy depends.⁷²

It might seem that this is the ideal first stage to begin an open-ended theory with. Unlike Fumerton's Hierarchy, it may not get any cases wrong. And unlike Sepielli's account, it can tell us *how* each kind of rationality depends on the agent's credences about a kind of value, and it tells us something about what this value is. However, it does have several limitations. The fact that it relies on the notion of maximizing expected value in order to explain these things makes it somewhat contentious. It could also make it a narrower normative theory. "Maximizing expected value," someone might hold, "is what *rationality* is all about. But other normative notions (perhaps the epistemic *ought*, for example) can't be explained in these terms."

The S&W Hierarchy was also not designed with cases of truth uncertainty about normative theories in mind. If 'value' is understood appropriately, then it can accommodate these cases as well. For example, if I am uncertain between two kinds of consequentialism, these might be translated into different theories of value. Then I could be said to be uncertain between two versions of level-1-rationality in the S&W Hierarchy. And perhaps it really would be most rational for me, given this uncertainty, to do what would maximize expected value.

However, if the account *is* meant to generalize to cases of truth uncertainty, then that only makes it more contentious. One competing view of how to deal with truth uncertainty about is the My Favorite Theory view, according to which we should not hedge our bets in

⁷² For example, what's rational'' depends on *the agent's credences about expected value* (of which expected value is a function), and what's rational' depends on *the agent's credences about value* (of which expected value is a function). Hence expected value does depend on 'the way the world is according to expressions of credences upon which rationality' depends,' since it simply depends on the credences upon which rationality' depends: that is, the agent's credences about value.

cases of truth uncertainty about normative theories like we do in cases of application uncertainty.⁷³ Instead, we are simply to try our best to do what is in fact required by the true normative theory. In terms of value maximization, this means that we should choose the option that's most likely to maximize value as opposed to the option that will maximize expected value. While the My Favorite Theory theory has some counterintuitive results, it also has defenders.

In the next section, I'll propose a hierarchy theory that attempts to find a happy medium between a plausible but very general account like Sepielli's, and a plausible but specific (and hence contentious and potentially too narrow) account like Spencer and Wells's.

3.4. Proposal for a Good First Stage: the *Lite* Hierarchy Theory

An open-ended normative theory is built in stages. The first stage answers all normative questions that don't involve uncertainty about any stages or the larger theory itself. The second stage answers all remaining questions that don't involve uncertainty about the second stage, later stages, or the larger theory itself. The third stage answers all remaining questions that don't involve uncertainty about the third stage, later stages, or the larger theory itself. And so on. Such a theory can always be extended, and is never meant to be complete.

This may sound just like a hierarchy theory. But in the hierarchy theories I've described, the 'stages' (or 'levels') are all unified by a single recursive definition. An open-ended theory will be more complex than this, and an infinite hierarchy theory can represent a single stage in it. However, hierarchy theories make for good first stages in open-ended theories because

⁷³ Gustafsson and Torpman 2014.

they cover so much normative space at once. My proposal for a first stage is a hierarchy theory that I'll call the '*Lite* Hierarchy Theory.'

To make the project manageable, I'll restrict its scope. First, I'll assume that the only normative questions to answer are about what agents *should* do. This isn't a significant change of topic from the hierarchy theories I've just discussed. For our purposes, the questions *what's it rational for one to do?* and *what should one do?* or *what should one do, given one's uncertainty about such-and-such?* could mean basically the same thing. I prefer *should* because I've used 'should' language in chapters 1 and 2.

Second, I'll assume that the only subjectivist normative questions to answer are those about what agents should do *given their uncertainty about something*. This is because these are the only subjectivist questions needed for the Incompleteness Argument posed in Chapter 2. In Appendix E I suggest ways in which the hierarchy theory I describe could be extended to answer other subjectivist questions as well.

In Chapter 1, I propose an account of should: 'ALL-THINGS-CONSIDERED SHOULD.' The account promises answers to every objectivist normative question about what agents should do, if supplemented with an account of the (relative) values of states of affairs. I'll use this notion of should as the first level in the hierarchy:

Level 1	<p>A OBJECTIVELY SHOULD ϕ</p> <p>= A ALL-THINGS-CONSIDERED SHOULD ϕ</p> <p>= If A were to ϕ, then one of A's most valuable reachable states of affairs would obtain.</p>
----------------	--

Each agent has a number of different options available to her. And for each of these options, it's true that, if she were to perform it, some set of states of affairs would obtain.

The set of all the states of affairs that would obtain if the agent were to perform one of her options or another is the set of the agent's *reachable* states of affairs.

Agent	A's options	States of affairs that would obtain conditional on A's performing one of her options		= A's <i>reachable</i> states of affairs
A	O1	→	S1	S1, S2, S3
	O2	→	S2	
	O3	→	S3	

Table 4. *Reachable states of affairs*

The ALL-THINGS-CONSIDERED SHOULD concept assumes that reachable states of affairs can be given an ordinal ranking according to their relative *values*. For example, suppose that we fill in some details as follows:

A's options	A's <i>reachable</i> states of affairs	From most to least valuable
<i>Walk to Work</i>	→ Arrive to work late	1. Arrive to work on time
<i>Walk to Work</i> <i>Barefoot</i>	→ Arrive to work late with sore feet	2. Arrive to work late
<i>Drive to work</i>	→ Arrive to work on time	3. Arrive to work late with sore feet

Table 5. *Reachable states of affairs ranked according to value*

In this case, A's most valuable reachable state of affairs is the state of affairs in which she arrives to work on time. It's a state of affairs that she has the ability to bring about or let happen—that is, it's reachable for her—and it's also the most valuable state of affairs of this

kind. So, according to my account of OBJECTIVE SHOULD, performing the option that would bring this state of affairs about (driving to work) is what A objectively should do.

But the first level isn't the important part of the theory I propose. We just need a starting place that's general enough to accommodate many substantive objective normative theories of should. And all substantive views about what objectively should be done can be expressed in value functions that rank states of affairs.⁷⁴

The most important part of a hierarchy theory is the description of how to build the next levels. To accommodate normative questions of the form *what should one do, given one's uncertainty about p?* I'll introduce another piece of terminology:

A's **certainly reachable** states of affairs: all states of affairs S such that a perfect reasoner who's only uncertain about what A is uncertain about (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

Your *certainly* reachable states of affairs are basically those that you can be certain of *how* to reach. To illustrate, consider Frank Jackson's 'Medicine Case':

Jill is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has three drugs to choose from: Drug A, Drug B, and Drug C. Careful consideration of the literature has led her to the following opinions. Drug A [will] relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other, though, will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which is the killer drug. What should Jill do?⁷⁵

⁷⁴ A simple (if imperfect) way to try to do this would be to give the state of affairs in which *one does what one should according to the theory* a value of 1 and all other states of affairs a value of 0.

⁷⁵ Jackson 1991 pages 462-3. I've changed the case slightly, from one where Jill has the opinion that Drug A is 'very likely' to relieve the skin condition to one where she has the opinion that it *will* relieve it. This difference actually matters in the theory I propose. While the theory can issue the same verdicts about both versions of the case, the story if Jill only thinks that Drug A is *very likely* to relieve the condition is more complex.

In this case, Jill's reachable states of affairs include: one in which John is completely cured, one in which John is partially cured, and one in which John is dead. If we suppose that Drug C will (in fact) completely cure the skin condition, then Jill's most valuable reachable state of affairs is the one in which John is completely cured. From this we can determine that she objectively should perform whichever one of her options would bring this about—in this case: administering Drug C.

But unfortunately Jill herself is not in a position to figure this out, since she doesn't know which of drugs B and C would completely cure the skin condition and which would kill her patient. She can determine that the state of affairs in which John is completely cured is *reachable*—that there is *something* she can do that would ensure that it obtains—but she doesn't know what that something is, exactly. If Jill were to fill in a table describing the states of affairs that would obtain conditional on her performing one of her options, she'd only get as far as this:

Jill's options	States of affairs that would obtain conditional on Jill's performing one of her options	Jill's reachable states of affairs
<i>Give Drug A</i>	→ John is partially cured	John is partially cured,
<i>Give Drug B</i>	→ ?	John is completely cured,
<i>Give Drug C</i>	→ ?	John is dead

Table 6. How Jill might complete a table showing her reachable states of affairs

So Jill's only *certainly reachable state of affairs* is the state of affairs in which John is partially cured.

It seems to many that, given Jill's uncertainty about what drugs B and C would do, Jill should administer Drug A. This can be explained by a single subjectivist theory of should according to which:

A SUBJECTIVELY SHOULD ϕ

= If A were to ϕ , then one of A's most valuable *certainly* reachable states of affairs would obtain.

With a theory like this, we can explain that Jill's only, and hence most valuable, certainly reachable state of affairs is the one in which John is partially cured. And that because of this, Jill subjectively should perform whichever of her options would ensure that this state of affairs would obtain—in this case: administering Drug A.

There are versions of Jackson's case about which the account may appear to give the wrong verdict, but does not. Consider:

Jill 2 is a physician who has to decide on the correct treatment for her patient, John, who has a serious infection. She has three drugs to choose from: Drug A2, Drug B2, and Drug C2. Careful consideration of the literature has led her to the following opinions. Drug A2 will completely fight off the infection for two months, but then John will die. One of drugs B2 and C2 will completely fight off the infection for only one month, at which point John will die; the other will completely cure him. There is no way for Jill 2 to tell which of B2 and C2 is the perfect cure. What should Jill 2 do?

Here's how Jill 2 might be able to fill in a table representing her reachable states of affairs:

Jill 2's options	States of affairs that would obtain conditional on Jill 2's performing one of her options	Jill 2's reachable states of affairs
<i>Give Drug A2</i>	→ John lives for 2 months	John lives for 2 months,
<i>Give Drug B2</i>	→ ?	John lives for 1 month,
<i>Give Drug C2</i>	→ ?	John is completely cured

Table 7. How Jill 2 might complete a table showing her reachable states of affairs

So does this mean that our theory must say that Jill 2 subjectively should give Drug A2?

While there are reasons to think this could be a good choice (it wouldn't put her patient at risk of immediate death, for one), there are also good reasons to think that in this case, giving one of the other drugs could be a better choice. Taking the risk in this second case could have a huge reward.

In making decisions, we usually focus our attention on the small handful of states of affairs that we take to be relevant. But in reality, we have a vast number of reachable states of affairs at any one time. For each of our options, many states of affairs would obtain if we were to perform it. So the '?'s in Table 7 aren't strictly speaking accurate. Jill 2 could find many states of affairs to add in there. Here's one of the more relevant ones:

Jill 2's options	States of affairs that would obtain conditional on Jill 2's performing one of her options	Jill 2's reachable states of affairs
<i>Give Drug A2</i>	→ John lives for 2 months	John lives for 2 months, John lives for 1 month, John is completely cured, There's a 50% chance that John lives for 1 month and a 50% chance that he's completely cured
<i>Give Drug B2</i>	→ There's a 50% chance that John lives for 1 month, and a 50% chance that he's completely cured	
<i>Give Drug C2</i>	→ There's a 50% chance that John lives for 1 month, and a 50% chance that he's completely cured	

Table 8. How Jill 2 might (more carefully) complete a table showing her reachable states of affairs

With this new certainly reachable state of affairs to consider,⁷⁶ it's no longer obvious that the state of affairs in which John lives for 2 months is the one that's most valuable. It could be that, according to the account of SUBJECTIVE SHOULD I've given, Jill 2 subjectively should give one of drugs B2 or C2.

However, there are three problems with my proposed account of subjective should. Each is a way in which the account is not subjective enough. The first is that ranking certainly reachable states of affairs according to their (objective) values may not quite work. For example, consider some of the other certainly reachable states of affairs Jill has in Jackson's original case:

Jill's options	States of affairs that would obtain conditional on Jill's performing one of her options	Jill's reachable states of affairs
<i>Give Drug A</i>	→ John is partially cured	John is partially cured,
<i>Give Drug B</i>	→ Jill gives John Drug B	John is completely cured,
<i>Give Drug C</i>	→ Jill gives John Drug C	John is dead,
		Jill gives John Drug B,
		Jill gives John Drug C,
		etc.

Table 9. How Jill might (more carefully) complete a table showing her reachable states of affairs

⁷⁶ Something like this state of affairs is needed here, but *the state of affairs in which there's a 50% chance that John lives for 1 month, and a 50% chance that he's completely cured* is probably not it. Note that this can't be a state of affairs in which the objective probability of John living for 1 month is 0.5 and the objective probability of John's being completely cured is 0.5. Once Jill gives Drug B, the objective probability may be near 1 that John lives for 1 month and near 0 that John is completely cured. (And *vice versa* if Jill gives Drug C.) But it also doesn't seem right to treat this as a state of affairs in which Jill has a subjective probability of 0.5 that John lives for 1 month and 0.5 that John is completely cured. That's because this is just the state of affairs in which Jill has some mental state. It's not clear why this would be assigned any more value than, e.g., the state of affairs in which John lives for 1 month. (Thanks to Andy Egan for this point.) So we need something different. One thing that could work is *the state of affairs in which Jill has performed an option that had the subjective probability of 0.5 of causing John to live for 1 month and the subjective probability of 0.5 of causing John to be completely cured*, or *the state of affairs in which Jill has performed the option with the highest expected utility*.

This table represents how Jill would be able to fill things in, given her uncertainty about what Drugs B and C do. So even though all the states of affairs on the right are reachable for Jill, only the smaller set on the left are certainly reachable. However, filling in these two new trivial states of affairs (*Jill gives John Drug B* and *Jill gives John Drug C*) causes problems.

According to the theory of subjective should I've proposed, Jill subjectively should do what's required to bring about one of her most valuable certainly reachable states of affairs.

Suppose that Drug C is in fact the cure. Then, arguably, the state of affairs in which Jill gives John Drug C is the most valuable of the three, since it will lead to John being completely cured. And the verdict would then be that Jill subjectively should administer Drug C. But that can't be right, since she has no idea that Drug C is the cure.

There are two possible views about how to assign (objective) value to a state of affairs like the one in which Jill gives John Drug C. On one view, let's call it the 'Wide View,' *the state of affairs in which Jill gives John Drug C* is as (objectively) valuable as *the state of affairs in which Jill gives John Drug C and John is cured*, since giving John Drug C will, as a matter of fact, cure him. On the other, 'Narrow View,' the first state of affairs is (objectively) less valuable than the second because it does not 'mention' or 'include' in some relevant sense, John's being cured.

I will not attempt to decide this issue here. But the Wide View would pose a problem for the account of subjective should I've given. And we can solve it by making the value function that my subjective account appeals to sensitive to the agent's uncertainty. So far, the account just uses the agent's uncertainty to determine which reachable states of affairs are relevant (i.e. the *certainly* reachable ones). These are then ranked according to an objective value function, which, at least on the Wide View, is what can allow objective facts to play too large a role in determining what subjectively should be done.

A subjective value function that takes into consideration what the agent is uncertain about can assign *the state of affairs in which Jill gives John Drug C* and *the state of affairs in which Jill gives John Drug C and John is cured* different (subjective) values, because it will be sensitive to the fact that Jill does not know that Drug C is the cure. And we can use this notion to fix up the account:

A SUBJECTIVELY SHOULD ϕ

= If A were to ϕ , then one of A's most *subjectively* valuable certainly reachable states of affairs would obtain.

With this distinction, we ensure the right result in Jackson's original case on either the Wide View or the Narrow View of objective value.

The second problem is that SUBJECTIVE SHOULD doesn't get the right result in cases where, unbeknownst to the agent, performing some other random option would bring about a most (subjectively) valuable certainly reachable state of affairs. For example, suppose we add a fourth mystery drug, Drug X, to the original case of Jill:

Jill 3 is a physician who has to decide on the correct treatment for her patient, John, who has a minor but not trivial skin complaint. She has four drugs to choose from: Drug A, Drug B, and Drug C, and Drug X. Careful consideration of the literature has led her to the following opinions. Drug A will relieve the condition but will not completely cure it. One of drugs B and C will completely cure the skin condition; the other, though, will kill the patient, and there is no way that she can tell which of the two is the perfect cure and which is the killer drug. Jill 3 has absolutely no idea what Drug X would do, but as it happens it would do exactly the same thing as Drug A. What should Jill 3 do?

Here's how Jill 3 could fill in the table:

Jill 3's options	States of affairs that would obtain conditional on Jill 3's performing one of her options	Jill 3's reachable states of affairs
<i>Give Drug A</i>	→ John is partially cured	John is partially cured,
<i>Give Drug B</i>	→ ?	John is completely cured,
<i>Give Drug C</i>	→ ?	John is dead
<i>Give Drug X</i>	→ ?	

Table 10. How Jill 3 might complete a table showing her reachable states of affairs

In this case, giving Drug A and giving Drug X *both* meet the conditions for options that Jill 3 subjectively should perform. Why? *John is partially cured* is still Jill 3's only certainly reachable state of affairs. And if she were to use Drug A, then it would obtain. But here it's also true that, if she were to use Drug X, this state of affairs would obtain. Jill 3 just doesn't know this. So according to our account, if Jill 3 were to give Drug X, she'd have acted as she subjectively should. But that's not right. It would be terribly reckless for Jill 3 to use the random drug, since she knows nothing about what it does.

We can make another adjustment to the theory to handle the case of Jill 3. It makes not only the set of relevant states of affairs and the value function that ranks them sensitive to the agent's uncertainty, but also the connection between the agent's options and the states of affairs that they would guarantee.

A SUBJECTIVELY SHOULD ϕ

= (i) If A were to ϕ , then one of A's most subjectively valuable certainly reachable states of affairs, S, would obtain, and (ii) ϕ is **certainly linked** to S.

In this version, an option, ϕ , is ‘certainly linked’ to a state of affairs, S , just in case a perfect reasoner who’s only uncertain about what A is uncertain about (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain. This theory gets the right result in the case of Jill 3. Since a perfect reasoner who’s only uncertain about what Jill 3 is uncertain about would not be able to determine that Drug X would partially cure John’s skin condition, the option of giving Drug X is not certainly linked to the state of affairs in which John is partially cured. Hence, it’s not true that Jill 3 subjectively should administer Drug X.

The final, and expected, problem with all these versions of the account of subjective should is that they cannot give good answers to every subjectivist should question of the form *what should one do given one’s uncertainty about such-and-such?* This won’t come as a surprise, given that the starting assumption of this chapter is that we will never be able to find a normative theory that can answer all these questions.

But we can easily extend the objective and subjective accounts of should into an infinite hierarchy theory that answers many more of these questions:

The *Lite* Hierarchy Theory

- | | |
|---------|--|
| Level 1 | <p>A OBJECTIVELY SHOULD ϕ</p> <p>= If A were to ϕ, then one of A’s most (objectively) valuable reachable states of affairs would obtain.</p> |
| Level 2 | <p>A SUBJECTIVELY₁ SHOULD ϕ</p> <p>= (i) If A were to ϕ, then one of A’s most subjectively₁ valuable certainly₁ reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly₁ linked to S.</p> |

Level 3	<p>A SUBJECTIVELY₂ SHOULD ϕ</p> <p>= (i) If A were to ϕ, then one of A's most subjectively₂ valuable certainly₂ reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly₂ linked to S.</p>
...	...

Where:

A's **certainly_n reachable states of affairs**: all states of affairs, S, such that, if A were only uncertain about what she subjectively_{n-1} should do and everything that forms the basis of her uncertainty about what she subjectively_{n-1} should do, a perfect reasoner who's only uncertain about these things (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

ϕ is, for A, **certainly_n linked** to S just in case: if A were only uncertain about what she subjectively_{n-1} should do and everything that forms the basis of her uncertainty about what she subjectively_{n-1} should do, a perfect reasoner who's only uncertain about these things (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

Subjective_n value: value assigned to states of affairs in a way that is appropriately sensitive to A's uncertainty about what she subjectively_{n-1} should do.

The *Lite* Hierarchy Theory works roughly as follows. You objectively should do what will ensure that the most valuable state of affairs you can possibly guarantee obtains. (If there were something you could have done that would have had a better result, then you objectively should have done that instead.) But you might be uncertain about how to do

this. If you are, then you subjectively₁ should ensure that the most valuable state of affairs obtains, of those that you are in the position to know *how* to guarantee—given your uncertainty about what you *objectively* should do, that is. But you might *also* not know how to do *this*. So in that case, you subjectively₂ should ensure that the most valuable state of affairs obtains, of those that you are in the position to know how to guarantee—given your uncertainty about what you *subjectively*₁ should do. And so on.⁷⁷

The theory looks more complicated than it is because of all the subscripts marking different degrees of sensitivity to uncertainty. While it's important to make these distinctions, it's also worth keeping in mind that the *Lite* Hierarchy Theory is just a more sophisticated version of the simple two-level theory:

Level 1	A OBJECTIVELY SHOULD ϕ = If A were to ϕ , then one of A's most valuable reachable states of affairs would obtain.
Level 2	A SUBJECTIVELY SHOULD ϕ = If A were to ϕ , then one of A's most valuable certainly reachable states of affairs would obtain.

To show (as opposed to merely claim) that more levels are needed, consider a fourth variation on Jackson's case:

⁷⁷ In this rough description of the view, your '*being in the position to know* how to guarantee a state of affairs, given your uncertainty about such-and-such' means something like: if you were a perfect reasoner who *only* has your uncertainty about such-and-such (and your basis for that uncertainty), then you would be able to determine with certainty how to guarantee that state of affairs. Note that while the subjective levels all have the phrase 'certainly reachable' in them, none of them depend on *your* actually being certain of anything.

Jill 4 is exactly like Jill 2, except that Jill 4 believes that the simple two-level account above is true, and is uncertain about how to use the account of subjective should. She knows that she should act so as to ensure that her most valuable certainly reachable state of affairs obtains. But she doesn't know whether the state of affairs in which John lives for 2 months is more or less valuable than the state of affairs in which there's a 50% chance that he lives for 1 month and a 50% chance that he is completely cured. What should Jill 4 do, given her uncertainty about *this*?

The simple two-level theory will say something about what Jill 4 subjectively should do. For example, it might say that she should give Drug A2. (If the state of affairs in which John lives for 2 months is really the more valuable one.) But it will not give an answer that's appropriately sensitive to what Jill 4 is uncertain about, since it will rely on the actual facts about which of Jill 4's certainly reachable states of affairs is most valuable.

3.5. Two Applications

An 'applications' section is usually a chance to show off how one's proposal solves some problems. And that is partly my intention here. I'll briefly describe two puzzling cases for normative theorists—the problem of moral uncertainty and the problem of higher-order defeat—and discuss how the *Lite* Hierarchy Theory could be used to make progress on them. But this section also serves to show the limitations of the *Lite* Hierarchy Theory; in particular, the fact that it's just a *first* stage in an open-ended theory.

3.5.1. The Problem of Moral Uncertainty

The problem of moral uncertainty is that we can be uncertain about which moral theory is true. And in these cases, the true moral theory may not provide the right answer to the question of what we should do, given this uncertainty. It raises some challenges:

1. How are we to make sense of the fact that what one should do when one has moral uncertainty doesn't always seem to be the same as what one should do according to the true moral theory?

And if we appeal to a different theory to tell us what to do in cases of moral uncertainty, then:

2. How are we to compare different moral theories when designing a theory that tells us what to do when one is uncertain between them?
3. How are we to handle the next obvious kind of uncertainty: uncertainty about the theory that tells us what to do when one is morally uncertain?

The *Lite* Hierarchy Theory offers a way to make some headway, primarily by providing a coherent way of thinking about what a theory that's designed to handle cases of moral uncertainty would be like.

Re 1: What you morally should do might be the same as what you objectively should do. But it can also be represented as a narrower kind of objective should that ranks states of affairs according to their *moral* values. Thus we'd get the following '*Lite* Moral Hierarchy Theory':

The *Lite* Moral Hierarchy Theory

Level 1	<p>A OBJECTIVELY MORALLY SHOULD ϕ</p> <p>= If A were to ϕ, then one of A's most (objectively) morally valuable reachable states of affairs would obtain.</p>
Level 2	<p>A SUBJECTIVELY₁ MORALLY SHOULD ϕ</p> <p>= (i) If A were to ϕ, then one of A's most subjectively₁ morally valuable certainly₁ reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly₁ linked to S.</p>
...	...

Uncertainty between different moral theories can be represented as uncertainty between different (objective) moral value rankings of states of affairs.

So, is the question about what to do when you are morally uncertain a moral question or not? This becomes a matter of terminology. We could call the whole hierarchy theory (or the larger open-ended theory to which it belongs) 'the moral theory,' we could call the first level 'the objective moral theory' and the next levels 'subjective moral theories' as I've done above, or we could call only the first level 'the moral theory' and the rest of the levels by different names entirely (e.g. theories of 'rationality'). The choice doesn't seem to matter much.

Re 2: Nothing in the *Lite* Moral Hierarchy Theory I've sketched says anything about how to express different moral theories as different moral value rankings of states of affairs, or about how to design a subjective value ranking that's sensitive to one's uncertainty about the objective moral value ranking. This is what the problem of moral uncertainty becomes if we adopt the *Lite* Hierarchy Theory. While it could prove helpful to have the problem framed in this way, we're still left with difficult work to do.

Re 3: Because it's a hierarchy theory, the *Lite* Moral Hierarchy Theory explains how to deal with the next kind of uncertainty (uncertainty about the theory that tells us what to do when we're morally uncertain), and the next, and so on. Each level can handle moral uncertainty about the previous level. But it should also be possible to be uncertain about the *Lite* Moral Hierarchy Theory as a whole. A theory that says what should be done given *this* kind of uncertainty would require another stage in the open-ended theory. I discuss next stages in section 3.6.

3.5.2. The Problem of Higher-Order Defeat

Suppose I believe that fresh coffee is nearby based on the smell, and that this belief is justified. Then I get good evidence that my belief is not justified—not (direct) evidence that there is no coffee nearby, but evidence that, say, experiencing the smell of coffee is a common side effect of the new medication I'm on. It seems that this 'higher-order' evidence about the justification of my belief causes my belief to be unjustified. If I were to continue to hold it and make plans to get some of the coffee, then I'd seem irrational. This is the phenomenon of higher-order defeat. Since it's best understood through examples, here are two more:⁷⁸

Mental Maths: My friend and I have often amused ourselves by solving little math problems in our heads, and comparing our answers. We have strikingly similar track records: we are both very reliable at doing mental maths, and neither is more reliable than the other. We now engage in this pastime, and I come up with an answer to a problem, 457. I then learn that my friend came up with a different answer, 459.⁷⁹

⁷⁸ These are based on two of David Christensen's most popular thought experiments illustrating the phenomenon, but described differently (and more compactly) in Lasonen-Aarnio 2014.

⁷⁹ Lasonen-Aarnio 2014 page 315; based on the Peer Disagreement case in Christensen 2010a pages 186-187.

Hypoxia: I have just achieved a difficult first ascent in the Himalayas. As the weather turns, I have to abseil down a long pitch. I have gone through a sequence of reasoning several times to check that I have constructed my anchor correctly, that I haven't underestimated the length of the pitch, and that I have threaded the rope correctly through my belay device and carabiner. I then acquire evidence that I am in serious danger of being affected by a mild case of hypoxia caused by high altitude. Such hypoxia impairs one's reasoning while making it seem perfectly fine. I know that mountaineers have made stupid but fatal mistakes in the past as a result of being in such a condition.⁸⁰

In both cases, it's supposed to be intuitive that I should give up a belief or lower my confidence in something on the basis of receiving evidence that my doxastic state is the product of a flawed process. Evidence about the rationality of a belief (or other doxastic state) is *higher-order evidence*. *Higher-order defeat* occurs when higher-order evidence removes—'defeats'—the rationality of a first-order belief.

The phenomenon of higher-order defeat poses a problem for accounts of epistemic rationality. Suppose our account of epistemic rationality looks like this:

It's (epistemically) rational for A to believe p *iff* C

Then, to explain higher-order defeat, we'd need to find a plausible set of conditions C such that evidence that C doesn't hold actually makes it so that C really doesn't hold. And this may not be possible to do. The trouble is that, whatever we fill in for C (for example, *A's belief that p is the product of a reliable process, p is supported by A's evidence*, etc.) it seems entirely possible to for one to have misleading evidence that C isn't met.⁸¹

The problem of higher-order defeat is a special case of the converse of Fumerton's Puzzle. His puzzle is that, for any account of rationality, it seems possible to have false but

⁸⁰ *Ibid*; based on the Pilot case in Christensen 2010b page 126.

⁸¹ Lasonen-Aarnio makes a compelling case for this in her 2014.

misleading evidence that some option is rational when it's not rational according to the account. And in these cases, we want to say that the agent who acts as she justifiedly believes she should act is rational, at least in some sense. The converse version is that: for any account of rationality, it seems possible to have false but misleading evidence that some options is *not* rational when it really *is* rational according to the account. And in these cases, we want to say that the agent who continues to act as she justifiedly believes she *shouldn't* is *not* rational (at least in some sense). Cases of higher-order defeat add the extra wrinkle that in these cases, the original rationality is supposed to be *removed* by the misleading (higher-order) evidence.

Fumerton solved his puzzle by distinguishing between different kinds of rationality. And the *Lite* Hierarchy Theory I've proposed offers us the same kind of solution, so long as we replace 'A should ϕ ' with 'it's rational for A to ϕ .'

The *Lite* Hierarchy Theory

Level 1	It's OBJECTIVELY RATIONAL for A to ϕ = If A were to ϕ , then one of A's most (objectively) valuable reachable states of affairs would obtain.
Level 2	It's SUBJECTIVELY ₁ RATIONAL for A to ϕ = (i) If A were to ϕ , then one of A's most subjectively ₁ valuable certainly ₁ reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly ₁ linked to S.
...	...

If there are an infinite number of kinds of rationality, then we can claim that cases of higher-order defeat really involve two kinds of (epistemic) rationality instead of one. So, for example, we might say that it's *objectively* rational for me to believe that there is fresh coffee

nearby. But once I get evidence that I may be smelling (or seeming to smell) coffee no matter what, it's not *subjectively 1* rational for me to continue to believe that there is coffee nearby.

While this kind of story could solve the converse of Fumerton's puzzle (or solve it as well as Fumerton's Hierarchy could solve the original puzzle), it's not quite enough here. We also need an explanation of higher-order defeat.

However, the explanation could be somewhat revisionary. It might turn out that higher-order defeat does not work quite like other kinds of defeat. That's because the problem of higher-order defeat only presents us with a few options. If it's always possible to have false but misleading evidence that a doxastic state isn't rational, then we could give up on ever finding a plausible account of epistemic rationality. But that seems like a last resort. We could deny that the phenomenon of higher-order defeat exists, and claim that it's actually perfectly rational for me to continue to believe that there is coffee nearby, or that the answer to the mental maths problem is 457, or that I've secured my rope correctly. That's definitely a *better* conclusion to draw. But it would be best if we could tell a different kind of story about cases of higher-order defeat without these counterintuitive consequences.

If there's a single kind of rationality, then higher-order defeat works to remove it. Some doxastic state is rational, then one gets defeating higher-order evidence, and the doxastic state is no longer rational. If we appeal to more than one kind of rationality, however, then 'defeating' higher-order evidence may have no effect on the original kind of rationality.

Consider the case Hypoxia. I begin with the rational belief that I have constructed my anchor correctly. Let's suppose it's an objectively (epistemically) rational belief. Were I to have this belief, one of my most valuable reachable epistemic states of affairs would obtain. For example, suppose that my only options are to give up this belief or to keep it, and that if

I were to give it up, this would result in an epistemically worse state of affairs: one in which I no longer have a true, reliably formed, well-founded, etc., belief about a matter of importance. Is this belief also subjectively₁ (epistemically) rational? Using the *Lite* Hierarchy Theory, this second level doesn't even apply to my case because I'm not yet uncertain about the first level. Once I receive evidence that I may be suffering from hypoxia, this higher-order evidence does cause me to be uncertain about what it's objectively rational for me to believe. On this picture, no rationality is *removed* by higher-order evidence. Instead, the higher-order evidence causes uncertainty about the rationality of the first-order belief, this uncertainty makes another kind of rationality applicable, and yet it doesn't apply because of the higher-order evidence.

Even a story like this provides some kind of solution to the problem of higher-order defeat. Sure, there is no actual *defeat* as no rationality is removed. But we don't have to say that it's perfectly rational for me to continue to hold my belief once I get higher-order evidence that it's irrational. We can admit that the belief remains perfectly rational in one sense, but we can also explain that the higher-order evidence has made the belief irrational in another sense. This may not look like defeat, exactly, but it does capture some of our intuitions about the cases.

However, there's also room to give an explanation that actually involves *defeat*. First, we could treat cases where there is no uncertainty about the objective level as ones where the subjective levels still apply, but trivially: they just say the same thing as the objective level. Then we could tell a different story, according to which I start out in the hypoxia case with a belief that's rational in *every* sense. The higher-order evidence and ensuing uncertainty make it so that my belief is no longer rational in all the subjective senses. On this way of

understanding things, higher-order evidence definitely removes the rationality of my belief—if not in *all* senses, then still in an infinite number of them!

A third strategy is to hold that there's always a privileged level in the hierarchy that's most important. To explain higher-order defeat, the privileged level would have to be the one that marks the 'highest' degree of the agent's uncertainty. In the hypoxia case, I would begin with an objectively rational belief, and no uncertainty about the first level in the hierarchy (or any others). Because of my lack of uncertainty, this first level would be the important one (even if all the other levels still apply). But then, with the evidence that I may be suffering from hypoxia, I would become uncertain about this first level. And so, the second level in the hierarchy would now be the most important one. And at that level, my belief would not be rational. On this way of understanding things, what higher-order defeat does is take one from being rational in the most important sense to being irrational in the most important sense, even though it may be that no kind of rationality is removed.

Finally, we could design a theory of should, or of rationality, that only cares about this important level:

The Important Level Theory

A should ϕ *iff* A should ϕ according to the important level of the *Lite* Hierarchy Theory.

(The 'important' level is the level applicable to the highest degree of uncertainty A has.)

This theory is worth mentioning because it's also a contender for a good first stage in our open-ended normative theory. It could fulfill almost the same role as the *Lite* Hierarchy Theory. And while the hierarchy theory is pluralist, this theory is monist. It also gives the best explanation of higher-order *defeat*. With the Important Level Theory, we can simply

explain that in cases of higher-order defeat: agents start out being rational according to the Important Level Theory, and then receive higher-order evidence that (by causing uncertainty) removes this very same kind of rationality.

The Important Level Theory has a drawback, though. It invites an extra kind of uncertainty: uncertainty about which level is the most important, or what one's highest level of uncertainty is. So it should be possible to design a case where my belief is rational for me to hold according to the Important Level Theory, but I receive misleading evidence that it's not. (For example, I may learn that I've taken a drug that causes most people to underestimate how much uncertainty they have.) Possibilities like this show that the Important Level Theory cannot explain all cases of higher-order defeat.

Perhaps this shows that the Important Level Theory is much worse than the *Lite* Hierarchy Theory? Allowing for an extra kind of uncertainty is a drawback. It means that the Important Level Theory leaves more subjectivist questions unanswered than the *Lite* Hierarchy Theory. But the *Lite* Hierarchy Theory doesn't entirely escape the problem either. Given the Incompleteness Argument, there will also be cases of uncertainty about the *Lite* Hierarchy Theory that it can't accommodate, and which could be used to generate similarly problematic cases of higher-order defeat.

The problems of moral uncertainty and higher-order defeat show three things about the *Lite* Hierarchy Theory. They show that it can provide some promising structure for building solutions, especially for making sense of the kinds of moral uncertainty and higher-order defeat that have been discussed so far (as opposed to those involving uncertainty about the *Lite* Hierarchy Theory itself). They also show that the *Lite* Hierarchy Theory leaves open many details about what it would actually take to be moral or (epistemically) rational.

Finally, they show that it leaves some cases of moral uncertainty and higher-order defeat with no solutions, and for which a *second* stage in the open-ended theory is required.

3.6. Next Stages

The *Lite* Hierarchy Theory might be a good first stage in an open-ended theory, but it can't be a complete theory. Like any normative theory, it can't accommodate cases of uncertainty about itself. But it's also not just like any normative theory. It's closer to a sketch of or framework for a normative theory. And it's because of this that it promises to handle most ordinary cases of uncertainty.

For example, are you uncertain about any objectivist normative theory that can be written in the form *A should ϕ iff C*? Your uncertainty can be expressed as uncertainty about the objective value ranking of states of affairs (uncertainty about what OBJECTIVELY SHOULD be done). Uncertainty about the *Lite* Hierarchy Theory itself will be more structural, or perhaps fundamental, than many cases of uncertainty about other normative theories.

How is it possible to be uncertain about the *Lite* Hierarchy Theory? For examples of truth uncertainty, we can look to its commitments. What is it committed to? All the definitions and biconditionals, but also to the existence of normative facts, subjunctive facts, privileged rankings of states of affairs according to their values, and an infinite number of subjectivist concepts. It's also possible to have application uncertainty about the *Lite* Hierarchy Theory. One might be uncertain about the nature of correct value rankings in a way that extends to every level. Or, one might be uncertain about some of the terminology used to define all the levels.

How could we build a next stage in our open-ended theory? In this section I'll discuss a very tempting strategy. Then I'll explain why I think we'll need to do something else—even though it's not obvious how to do it.

We could design a second stage that is very different from the first. But if we think that the *Lite* Hierarchy Theory works as a first stage, then we must think it handles cases of uncertainty correctly.⁸² So if we find ourselves needing some extra theory to say what's to be done when someone is uncertain of the *Lite* Hierarchy Theory, it would make sense to add another stage that looks almost exactly like the *Lite* Hierarchy Theory. For example, consider:

The *Lite* Hierarchy Theory 2

Level 1	<p>A SUBJECTIVELY_i SHOULD ϕ</p> <p>= (i) If A were to ϕ, then one of A's most subjectively_i valuable certainly_i reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly_i linked to S.</p>
Level 2	<p>A SUBJECTIVELY_{ii} SHOULD ϕ</p> <p>= (i) If A were to ϕ, then one of A's most subjectively_{ii} valuable certainly_{ii} reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly_{ii} linked to S.</p>
...	...

A's **certainly_i reachable states of affairs**: all states of affairs, S, such that, if A were only uncertain about the *Lite* Hierarchy Theory and everything that forms the basis of her uncertainty about the *Lite* Hierarchy Theory, a perfect reasoner who's only

⁸² Note that even if we think that, say, moral uncertainty needs to be treated differently than other kinds of uncertainty, this can be accommodated by the *Lite* Hierarchy Theory with appropriate subjective value rankings.

uncertain about these things (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

This second hierarchy theory would start with a new kind of subjective should that works in exactly the same way as the subjective shoulds of the first hierarchy, except that it involves uncertainty about the *Lite* Hierarchy Theory itself as opposed to uncertainty about any level in the *Lite* Hierarchy Theory.

If this strategy could work, then it would also be easy to see how the rest of the stages in the open-ended theory could be filled in. Would it threaten to unify the open-ended theory too much? If we attempt to define our open-ended theory as an infinite hierarchy of hierarchies like the *Lite* Hierarchy Theory, the *Lite* Hierarchy Theory 2, and so on, then it will be possible to be uncertain about the entire open-ended theory. (Which would also no longer be ‘open-ended’ at all). So that kind of definition could not work. But the idea that we could keep adding in this way forever *might* work, so long as the open-ended theory itself always remains somewhat incomprehensible to us. We can call the result of extending the *Lite* Hierarchy Theories forever in a true open-ended fashion a ‘Super Hierarchy Theory.’ Such a theory would not just be a single hierarchy or even a hierarchy of hierarchies, but would involve hierarchies of hierarchies forever. However, it would have to be designed so that if we ever were to try to be uncertain about it, we’d only succeed in being uncertain about some smaller portion of it. There’d always be another *Lite* Hierarchy Theory to give answers to what should be done given our uncertainty.

Unfortunately, we don’t have to go this far to see that this strategy won’t work. The *Lite* Hierarchy Theory 2 can’t be the next stage in the open-ended theory. This is because, as I’ve argued in Chapter 2, a theory can only give an adequate answer to a question of the form

what should one do given one's uncertainty about p? if it doesn't assume or *rely on* the truth about whether p in giving its answer. Consider the following case of uncertainty about the *Lite* Hierarchy Theory:

Jill 5 is just like Jill in Jackson's original Medicine Case. She has three options: give Drug A, give Drug B, and give Drug C. She knows that Drug A would partially cure her patient's skin condition, and that one of drugs B and C would completely cure it while one would kill her patient. She's trying to decide what to do, and she is unsure whether she should do what will maximize expected utility or do what will maximize the chance of maximizing utility. According to the first view, she should give Drug A. But according to the second, she should give one of drugs B or C. She has also adopted the *Lite* Hierarchy Theory, and represents her uncertainty as uncertainty between two ways of assigning subjective values to states of affairs: one where states of affairs that maximize expected utility are always the most subjectively valuable, and one where states of affairs that maximize the chance of doing what one objectively should are always the most subjectively valuable. She thinks of each as a version of the *Lite* Hierarchy Theory, 'Version 1' and 'Version 2,' and is 40% confident in the first and 60% confident in the second. This leads her to be very confused about what she should do. At every level of each hierarchy, the two versions disagree: Version 1 keeps telling her to give Drug A, while Version 2 keeps telling her to give one of the other drugs. What should Jill 5 do, given her uncertainty about the *Lite* Hierarchy Theory?

None of the levels in the *Lite* Hierarchy Theory 2 can (adequately) answer this question.

That's because each relies on a subjective value ranking. This value ranking will either

involve maximizing expected utility, or maximizing the chance of maximizing utility, or something else: it will not be neutral about what Jill 5 is uncertain about.⁸³

There is, however, a response to be made on behalf of the proponent of the *Lite* Hierarchy Theory 2. The requirement that an answer to a question of the form *what should one do given one's uncertainty about p?* not rely on or presuppose the truth of whether *p* might have *some* limitations. Consider the case of someone who's uncertain about the *Lite* Hierarchy Theory because she's uncertain about the existence of normative truths. *Any* possible next stage in an open-ended normative theory will have to presuppose the existence of normative truths. And that seems okay. Perhaps, following Jacob Ross, if you're uncertain whether a normative theory is true because you have some credence in normative nihilism, then your uncertainty can just be ignored.⁸⁴ At the very least, one might think, it could be fine to use the first level in the *Lite* Hierarchy Theory 2 to cover normative facts about what one should do given this kind of uncertainty.

I'll admit that uncertainty about the existence of normative facts is a special case. And this *might* mean that some other 'radical' or 'fundamental' kinds of uncertainty could also be given special treatment. Perhaps, for example, it would be going too far to require the next stage in an open-ended theory to rely on something other than subjunctive relations just to accommodate cases of uncertainty about them. But the case of Jill 5 is different. There, it does seem that we want a more neutral answer to the question of what Jill 5 should do than any that could be given by a next stage like the *Lite* Hierarchy Theory 2.

⁸³ It could be that Jill 5 is only uncertain about subjective value rankings in the *Lite* Hierarchy Theory. In that case, the *Lite* Hierarchy Theory 2 could be neutral about what she is uncertain about. But I'm imagining that Jill 5 is uncertain about subjective value rankings of states of affairs, period. It's also worth noting that if this kind of uncertainty is possible, which it seems to be, then no 'Super Hierarchy Theory' could be made to work since it would be possible to be uncertain about the whole thing.

⁸⁴ See, e.g., Ross 2006a.

In this chapter, I've argued that even if we could never hope to find an ultimate normative theory, we can build an open-ended one instead. And I've proposed the *Lite* Hierarchy Theory as a good first stage, or sketch of a good first stage, in an open-ended normative theory. The theory provides a framework that could be used to give adequate answers to a huge number of subjectivist normative questions, and to accurately cover a vast portion of normative space. It's designed with parts left to be filled in (notably, the value rankings), but it's more informative than Sepielli's sketch of a hierarchy theory, and less controversial than Spencer and Wells's hierarchy theory. It's not obvious to how to build a stage to follow it, but this may be an indication of two things: that the *Lite* Hierarchy Theory really is a *very* good first stage (since the only subjectivist questions it leaves unanswered are so difficult to know how to answer), and that there is more interesting normative theorizing to be done.

CONCLUSION

I began this dissertation with the assumption that the final goal of normative theorizing is to find or build an ‘ultimate’ normative theory—a theory that covers or ‘maps’ all normative facts, or that (adequately) answers all normative questions. And I’ve discussed two main obstacles that stand between us and this goal.

The first obstacle is that of apparent conflict between different normative domains. If we can’t tell a coherent story about how different normative domains interact, or if our best story fails to capture our intuitions about these cases, then it may not be possible to find an ultimate normative theory. An ultimate normative theory, after all, would not be incoherent, and it would be unlikely to be highly counterintuitive.

The second obstacle is that of finding a normative theory that could accommodate all subjective normative facts. An ultimate normative theory would cover all normative space, so if it’s impossible to find a theory that can do this, then it’s impossible to find an ultimate normative theory.

The final (positive) product of this dissertation is the *Lite* Hierarchy Theory, proposed as a first stage in an open-ended normative theory. If we adopt it, how do we stand?

The *Lite* Hierarchy Theory uses the Normativity *Lite* Framework, and the first level in the hierarchy is the all-things-considered account of should proposed in Chapter 1. Because of this, it promises a good story about apparent conflict between different normative domains. There are two options here. If there is a single privileged ranking of states of affairs according to their (objective) values, then there will be a single all-things-considered should concept that unifies all objective normative facts. This is the simplest way to think of the first level in the *Lite* Hierarchy Theory. But it’s also possible to use the *Lite* Hierarchy

Theory even if there isn't a single privileged ranking of states of affairs according to their (objective) values. In that case, there would be a number of different all-things-considered should concepts (each with a different way of ranking states of affairs), and each would form the basis of its own hierarchy of levels along the lines of the *Lite* Hierarchy Theory. 'The *Lite* Hierarchy Theory' would then refer to the conjunction of all of these hierarchies.

The rest of the levels in the *Lite* Hierarchy Theory are designed to accommodate subjective normative facts. However, as I argue in Chapter 2, no theory could accommodate them all. This means that the goal of finding an ultimate normative theory is an impossible one to reach. That's why the *Lite* Hierarchy Theory is only supposed to be a first stage in an open-ended normative theory—a theory that's meant to be extended indefinitely. In building a theory like this, we can still in some sense aim for an ultimate normative theory. We can continue to improve and expand our normative theory while recognizing that it's not a task we could ever complete.

APPENDIX A.

Adjusting the Normativity *Lite* Schema

The Normativity *Lite* Schema expresses the underlying similarity between all option-normative concepts. By directly applying the schema, we'd get *lite* versions of common option-normative concepts like:

- | | |
|--|--|
| 1. A SHOULD _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 2. A OUGHT TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 3. A MUST _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 4. A IS REQUIRED TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 5. A IS OBLIGATED TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 6. It's OBLIGATORY _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 7. A IS PERMITTED TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 8. It's PERMISSIBLE _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 9. A HAS A REASON TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 10. A HAS REASON TO _{lite} ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 11. It's REASONABLE _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 12. It's RATIONAL _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 13. It's MORAL _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 14. It's PRUDENT _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 15. It's GOOD _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 16. It's ADVISABLE _{lite} for A to ϕ | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 17. A IS JUSTIFIED IN _{lite} ϕ ing | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 18. A's ϕ ing is WARRANTED _{lite} | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |
| 19. A IS PRAISEWORTHY _{lite} for ϕ ing | (relative to state of affairs S) <i>iff</i> if A were to ϕ , then S |

The problem is that there are important differences between at least some of these concepts, and they're all given exactly the same analysis. The remedy is to make some small adjustments to capture the differences without straying too far from the schema.

Requirements

In Chapter 1, I explained how to adjust the schema to capture an important feature of requirement-like concepts like 1-6. So, for instance, (1) becomes:

- 1*. A SHOULD_{lite} ϕ (relative to state of affairs S) *iff* (**ϕ ing is the maximally disjunctive option such that:**) if A were to ϕ , then S

But this one qualification doesn't get us everything we want. For example, it might seem that if you should do one thing, and in order to do it you'd need to do another thing, then you're required to do that second thing as well. Suppose a doctor should give some medicine to his patient, and can't do this without filling a syringe with the medicine. Then it also seems that he should fill a syringe with the medicine. On the revised definition given, if he should_{lite} give his patient the medicine (relative to the state of affairs in which his patient's life is saved), and he can't do this without filling a syringe with the medicine, it does *not* follow that he should_{lite} fill a syringe with the medicine (relative to the state of affairs in which his patient's life is saved). This could be because it's not true that *if he were to fill a syringe with the medicine, the patient's life would be saved*. (For example, the doctor might fill the syringe but then give it to someone else or empty it onto the floor.)

But we can find a concept that behaves in this way. For example, we can distinguish between 'full' and 'partial' requirements, where options that are necessary to performing a full requirement are 'partially' required.

- 1a*. A SHOULD_{lite}^{FULL} ϕ (relative to state of affairs S) *iff* (ϕ ing is the maximally disjunctive option such that:) if A were to ϕ , then S
- 1b*. A SHOULD_{lite}^{PARTIAL} ϕ (relative to state of affairs S) *iff* (i) A SHOULD_{lite}^{FULL} ψ (relative to state of affairs S), and (ii) one must ϕ in order to ψ .

Making these adjustments could distinguish requirement-like concepts from the others on the list, but would not distinguish between the first six. A revisionary claim would be that we can make do with a single requirement-like concept here instead of six (or with two kinds of this concept: a ‘full’ and a ‘partial’ one). But further distinctions could be made between different requirement-like concepts if there’s a good enough reason to do so.⁸⁵

Permissions

With a direct application of the *lite* schema, it’s only permissible_{lite} for me to do something (relative to some state of affairs) if that state of affairs would obtain were I to do it. For example, it’s only permissible_{lite} for me to eat an ice cream sundae (relative to the state of affairs in which I don’t get heart disease) if it’s the case that: if I were to eat the sundae, I wouldn’t get heart disease. Or, suppose that a doctor’s patient is not going to recover, no matter what. Then it’s not permissible_{lite} for the doctor to spend more time with her (relative to the state of affairs in which she recovers). This analysis of permissibility seems too strict. It would prevent many options from being permitted_{lite} when it seems that they should be permitted. One remedy is to opt for a very weak notion of permissibility_{lite}, such that an option is permitted_{lite} relative to state of affairs S just in case performing it wouldn’t preclude S’s obtaining:

⁸⁵ For example, if there’s a difference between strong and weak necessity modals like ‘must’ or ‘have to’ versus ‘should,’ then we may want to preserve it. One way of doing so would be to define a notion of ‘strong requirement’ like:

1c*. A is STRONGLY REQUIRED_{lite} to ϕ (relative to state of affairs S) *iff* if A were to ϕ then S, and if A were not to ϕ then not S.

On this account, you’re strongly required_{lite} to do something relative a state of affairs just in case your doing it is a necessary as well as a sufficient condition for that state of affairs to obtain.

- 7a*. A is (WEAKLY) PERMITTED_{lite} to ϕ (relative to state of affairs S) *iff* **it's not the case that**, if A were to ϕ , then **not** S

Another option is to rely on likelihoods and thresholds. So, for example, we might go with:

- 7b*. A is PERMITTED_{lite} to ϕ (relative to state of affairs S) *iff* if A were to ϕ , then **it would be (or would remain) sufficiently likely that S**

There are other possibilities as well, but both of these would solve the problem.

Reasons

An important connection between requirements and reasons seems to be that if you *should* ϕ then there's *a reason* for you to ϕ , but it doesn't follow from the fact that there's *a reason* for you to ϕ , that you *should* ϕ . One way to preserve this is to make reasons facts that make requirements sufficiently likely:

- 9a*. R is A REASON_{lite} for A to ϕ (relative to state of affairs S) *iff* R makes it sufficiently likely that A SHOULD_{lite} ϕ (relative to state of affairs S)⁸⁶

Or, we might borrow from Stephen Finlay's account of reasons, according to which they are, roughly, *explanations* for why someone ought to do something:⁸⁷

- 9b*. R is A REASON_{lite} for A to ϕ (relative to state of affairs S) *iff* R is an explanation of why A SHOULD_{lite} ϕ (relative to state of affairs S)⁸⁸

⁸⁶ Or, if we distinguish between full and partial requirements:

- 9a*. R is A REASON_{lite} for A to ϕ (relative to state of affairs S) *iff* either: (i) R makes it sufficiently likely that A SHOULD_{lite}^{FULL} ϕ (relative to state of affairs S), or: (ii) R makes it sufficiently likely that A SHOULD_{lite}^{PARTIAL} ϕ (relative to state of affairs S).

⁸⁷ Finlay 2014 Chapter 4.

⁸⁸ Or, if we distinguish between full and partial requirements:

On both accounts, if you should_{lite} do something relative to some state of affairs S, then there is a reason_{lite} for you to do it (relative to S). Suppose that you should_{lite} drive Darcy to her job interview relative to the state of affairs in which she gets there on time. Let's suppose that should_{lite} is understood in our revised sense so that driving Darcy is your maximally disjunctive option such that, if you were to do it, she would get to her interview on time.⁸⁹ If you should_{lite} drive Darcy to her job interview relative to the state of affairs in which she gets there on time, there are presumably facts that make this likely as well as facts that explain it. One fact that does both is: *that Darcy's job interview is in 30 minutes, the driving time is 25 minutes, and no one else but you can drive her.*

I'm unsure whether this is just *one* fact rather than three or more facts, or, more importantly, whether each part of this fact doesn't count as a reason on its own. If we want to include, say, the fact *that no one else can drive Darcy* as a reason_{lite} for you to drive her (relative to the state of affairs in which she gets to her interview on time), then we'll have to fiddle with these definitions.⁹⁰ I'm also not confident that either of the proposed definitions are adequate analyses of the concept REASON_{lite}. Both rely on notions that call for their own definitions: 'sufficiently likely' and 'explanation.' And we'll also want to individuate reasons in a way that allows us to weigh them, as well as to explain the difference between one's *having* a reason and there *being* a reason for one to do something.

9b*. R is A REASON_{lite} for A to ϕ (relative to state of affairs S) *if and only if* either: (i) R is an explanation for why A SHOULD_{lite}^{FULL} ϕ (relative to state of affairs S) or (ii) R is an explanation for why A SHOULD_{lite}^{PARTIAL} ϕ (relative to state of affairs S)

⁸⁹ So it's not the case that if you were to *call her a cab* or *let someone else drive her*, she'd get there on time. And the truth of this should_{lite}-fact also requires that it's impossible that you, say, drive Darcy but get a flat tire or become locked in traffic and fail to get her to the interview on time.

⁹⁰ For example, we might say that a reason_{lite} is a *necessary* part of an explanation for a requirement, or a *necessary* part of a fact that makes the requirement likely.

However, my aim in this section is just to offer examples of the kinds of adjustments that are still open to us if we adopt the Normativity *Lite* Schema. And the schema is compatible with a wide variety of different accounts of option-normative concepts.

APPENDIX B.

Relationalism and Finlay's End-Relationalism

Normative relationalists hold that there's no such thing as 'unqualified' normative concepts. For example, there's no such thing as what you should *simpliciter* to do. Instead, there might be what you should *given your goals* to do, or what you should *relative to a set of rules* to do.⁹¹ According to them, normative concepts involve a relation to some further interesting kind of thing beyond agents and their options.⁹² The extra interesting kind of thing might be one of the agent's goals, a set of cultural beliefs, a set of standards, rules, and so on.

The most thoroughly developed relationalist view is Stephen Finlay's 'end-relationalism.' The theory is a version of what I'll call 'neo-instrumentalism.' Neo-instrumentalists are inspired by how intelligible and naturalist-friendly instrumental normativity is. Of all the kinds of normativity, that involved in facts like *she should wear a sweater in order to stay warm*, or *he mustn't water his cactus so much if he wants it to grow*, has been thought the easiest to explain. A popular way of doing this is to posit a means-end relation between the agent's option and her goal, and to maintain that this relation plus the fact that *the end is something the agent wants* make the relation normative. Thus we have:

Simple Instrumentalism

A SHOULD ϕ *iff* ϕ ing is a means to one of A's ends.

⁹¹ Relationalism could simply be called 'relativism,' but since it's much broader than the kinds of relativism we're familiar with, it's useful to have a different name for it.

⁹² Some examples of uninteresting kinds of things: times, worlds, circumstances, etc. where these are supposed to be the actual ones that the agent is in or at.

The neo-instrumentalist wants to explain *all* normativity with this means-end relation—not just instrumental normativity. And for other kinds of normativity, what the agent wants is irrelevant. For example, it might be morally impermissible for me to eat animals, however much I enjoy the taste. And my friend may be epistemically irrational for refusing to believe in climate change, even though he'd rather hold a more optimistic view.

To remedy this, the neo-instrumentalist becomes a relationalist and *removes the requirement that the end be anyone's goal*. On this view, a fact is normative just in case it involves a means-end relation between an agent's option and some 'end'—which is now represented as any state of affairs at all.

Neo-Instrumentalism

A SHOULD ϕ (relative to a state of affairs S) *iff* ϕ ing is a means to S.

Finlay's version of neo-instrumentalism provides a probabilistic analysis of the means-end relation. It's also primarily a descriptive theory of normative language. The central thesis is that all normative language has an end-relational semantics: that the semantics of any normative utterance involves relation to an end (a state of affairs). The semantic values of every first-order normative statement—for example 'chocolate is good,' 'there's no reason not to get a second opinion,' 'he should take his umbrella,' 'you ought not murder,' etc.—are said to essentially involve a relation to some contextually-supplied end. For Finlay, this relation is a probabilistic one, and the semantics of any normative utterance works in roughly the same way. Let's look at his account of the normative use of 'ought.'⁹³ Finlay's proposal is that the semantic content of an utterance of 'S ought to ϕ ' is:

⁹³ An example of a non-normative use of 'ought' is 'it ought to rain tomorrow,' where this simply means something like: it's likely to rain tomorrow.

‘A ought to ϕ ’: $\text{pr}(A \phi \text{ s} | e \ \& \ b) > \text{pr}(A \psi \text{ s} | e \ \& \ b)$, for all ψ such that $A \psi \text{ s}$ is a relevant alternative to $A \phi \text{ s}$.⁹⁴

Let’s work toward this step-by-step. First and foremost, the theory is relationalist, meaning (for a theory of language) that all normative statements must be understood in relation to *some other thing*. In Finlay’s case, this *other thing* is an ‘end,’ by which he just means a state of affairs. For example, even though you might simply say ‘white flour is best’ or ‘you ought to use white flour,’ according to Finlay these utterances only make sense relative to, say, the goal of feeding one’s sourdough starter. Luckily, our normative utterances do make sense on Finlay’s view because the context of each normative utterance supplies such an end.

Step 1 ‘A ought to ϕ ’: A ought to ϕ relative to e .

But positing contextually-determined ends doesn’t yet tell us much about their role or about the semantic content of ‘ought.’ According to Finlay, ‘ought’-statements make claims about the conditional likelihood of certain states of affairs. Very roughly, to say that ‘you ought to take Highway 9’ (in the appropriate context) is to say that you can be expected to get to your destination if you take Highway 9.

Step 2 ‘A ought to ϕ ’: A’s ϕ ing makes it likely that e .

Finlay understands probability as a measure of a possibility-space or “a set of possibilities \mathcal{W}_b defined by consistency with a partial description of a world-state which can be identified with the background b .”⁹⁵ On his view, the context of utterance also delivers a background,

⁹⁴ Finlay 2014 page 73.

⁹⁵ *Ibid.* page 44.

b , which fixes not only a possibility-space, but a unique probability distribution over propositions.⁹⁶ Adding the background-determined possibility-space W_b gets us:

Step 3 ‘A ought to ϕ ’: A’s ϕ ing makes it likely (relative to W_b) that e .

But Finlay thinks that ‘ought’-statements of the form ‘A ought to ϕ ’ are really *contrastive* probability statements. They don’t simply say that A’s ϕ ing will make an end likely; they say that A’s ϕ ing will make the end *more* likely than if A were to do something else. And not just anything else: only things that are relevant alternatives to A’s ϕ ing, given the context.

Step 4 ‘A ought to ϕ ’: A’s ϕ ing makes it more likely (relative to W_b) that e than A’s ψ ing, for all ψ such that $\mathcal{A}\psi$ is a relevant alternative to $\mathcal{A}\phi$.

This is basically Finlay’s final semantics for ‘S ought to ϕ ’ in less formal language.⁹⁷ His end-relational semantics is contextualist, and the context of utterance does a lot of work, fixing not only an end but also a background, probability function, and set of relevant alternatives.

To compare Finlay’s view with the Normativity *Lite* Framework, let’s pretend that Finlay’s view is not about language but about concepts:

⁹⁶ *Ibid.*

⁹⁷ However, in formalizing he does make choices worth noting. First, he switches from talk of the likelihood of e given what A does ($\text{pr}(e | \mathcal{A}\phi)$) to talk of the likelihood of what A does given e : $\text{pr}(\mathcal{A}\phi | e)$. This might seem problematic, as it’s possible to imagine an agent who’s really unlikely to choose to ϕ , even if ϕ ing is clearly the option that’s most likely to lead to e . (For example, the relevant end might be arriving at the cabin before noon, and A’s taking path 1 might be the action that’s most likely to achieve this end. But path 1 is snake-infested, and A is very afraid of snakes. A is physically able to take that route, but won’t unless she really needs to.) However, on Finlay’s theory, backgrounds don’t include information about the relative likelihoods of the agent’s taking each action: “[r]elative to such a background, each alternative has equal probability.” (Finlay 2014 page 73.) Second, he makes the probability of the agent’s actions conditional not on the end, which is simply a state of affairs, but on the end plus the background. This narrows the focus from *all* possible world-states where the end obtains to ones where the background conditions also hold. Such background conditions might introduce constraints like: *our actual physical laws hold, there is no alien invasion or unrelated catastrophic event*, etc. (Finlay 2014 page 40.)

Finlay's End-Relationalist Schema (for SHOULD; conceptual version)

A SHOULD ϕ (relative to end e) *iff* A's ϕ ing makes it more likely that e than A's ψ ing
(for all ψ such that $\mathcal{A} \psi$ is a relevant alternative to $\mathcal{A} \phi$)

This is very similar to the Normativity *Lite* Schema. Finlay's full end-relational view has quite a few differences from the Normativity *Lite* Framework. It's about normative language as opposed to normative concepts, and it's supposed to analyze all normative terms, not just the 'option-normative' ones. And it's a descriptivist project as opposed to a potentially revisionary one, which rules out distinguishing between minimal or '*lite*' normative concepts and using them to find or build more complex (and 'full-fledged') normative concepts like my all-things-considered should concept.

But why not build the Normativity *Lite* Framework using a probabilistic schema for SHOULD instead of a subjunctive one? First, note that we can express any one of Finlay end-relational should concepts using the Normativity *Lite* Schema. If we specify a particular end, e , we can get:

A SHOULD_{*lite*} ϕ (relative to S) *iff* if A were to ϕ , then: A would have done something that makes it more likely that e than any of A's other options (that are relevant alternatives)

Second, note that Finlay's schema is supposed to be neutral between many different first-order normative theories. For example:

A SHOULD ϕ (relative to state of affairs in which *utility is maximized*) *iff* A's ϕ ing makes it more likely that *utility is maximized* than A's ψ ing (for all ψ such that $\mathcal{A} \psi$ is a relevant alternative to $\mathcal{A} \phi$)

A SHOULD ϕ (relative to state of affairs in which *expected utility is maximized*) *iff* A's ϕ ing makes it more likely that *expected utility is maximized* than A's ψ ing (for all ψ such that $\mathcal{A} \psi$ is a relevant alternative to $\mathcal{A} \phi$)

A SHOULD ϕ (relative to state of affairs in which *A does what the virtuous person would do*)
 iff A's ϕ ing makes it more likely that *A does what the virtuous person would do*
 than A's ψ ing (for all ψ such that *A ψ s* is a relevant alternative to *A ϕ s*)

A SHOULD ϕ (relative to state of affairs in which *A believes in accordance with the evidence*)
 iff A's ϕ ing makes it more likely that *A believes in accordance with the evidence* than
 A's ψ ing (for all ψ such that *A ψ s* is a relevant alternative to *A ϕ s*)

But while Finlay's schema is entirely neutral between ends or states of affairs (just like the *lite* schema), it also adds in a bunch of machinery that's not neutral. Note that none of these end-relational concepts above would be entirely accurate ways of representing utilitarianism, expected utility theory, virtue theory, or evidentialism. The probabilistic relation adds something extra, so that it's no longer the case that *doing what maximizes utility*, or *acting virtuously*, or *believing in accordance with the evidence* is what should be done; instead, one should do what makes these things *likely*.

The *lite* schema doesn't add in anything like this. It could simply say, for example, that:

A SHOULD_{lite} ϕ (relative to the state of affairs in which *A believes according to the evidence*)
 iff if A were to ϕ , then: *A believes according to the evidence*.

This *lite* concept retains the evidentialist idea that believing according to the evidence is both necessary and sufficient for acting as one (epistemically) should.

So the subjunctive relation seems to bring fewer commitments than a probabilistic one. Since states of affairs can do so much work themselves (e.g. finding the right state of affairs lets us express any of Finlay's complex end-relational notions with the *lite* schema), the Normativity *Lite* Schema lets them do *all* the work, relying only on the more theoretically 'light-weight' subjunctive relation.

APPENDIX C.

Uncertainty

The Incompleteness Argument relies on a notion of uncertainty that I don't say much about. Because I think the argument could work with a number of different views, I don't want to commit myself to a contentious one. But I do have something in mind.

By 'uncertainty about p ' I mean 'felt' uncertainty about p ,⁹⁸ and in particular: *the inability or unwillingness to assume p for the sake of making some decision*. So uncertainty about p , or between p and not- p , need not be a 50% confidence. It might be that I'm, say, 70% sure that I locked my door, but I'm trying to decide whether I should go back home to check. In making that decision, I won't be willing to assume that I locked my door. My less-than-complete level of confidence matters to me, even if it's not that low.

Felt uncertainty about p is also not simply 'having a credence in p of less than 1.' But 'having a credence less than 1' is a more ordinary way of thinking about uncertainty. And if that's what's meant, then we are always uncertain about almost everything.

I want to remain neutral about whether this kind of uncertainty requires special treatment by normative theories. Suppose I'm trying to decide whether I should go home to check that I locked my door. Strictly speaking, I may have a credence of less than 1 that I still have a home, that I live where I think I live, and so on. But for the purpose of the decision I'm making, I'm assuming all of these things.

So what of the question *what should I do, given my uncertainty about whether I still have a home?* Must it be answered by a normative theory that doesn't rely on the truth about whether I do

⁹⁸ What I mean by 'felt' uncertainty is similar to what Sepielli means by 'conscious' (as opposed to 'dispositional') uncertainty in his 2014 page 91.

or don't? I'm not sure. It might require such an answer, and just not be an interesting question to ask. But the first kind of uncertainty can generate the Uncertainty Problem on its own.

The *Lite* Hierarchy Theory, on the other hand, will have to be designed more carefully with both kinds of uncertainty in mind. That's because many of the definitions involve *what a perfect reasoner, who's only uncertain about a level in the hierarchy in the way that some agent is (including the agent's basis for her uncertainty), could be able to be certain of*. I could stipulate that it's only the felt uncertainty about levels that matters. (This may be necessary to get the Important Level Theory to work, for example, since there may be a highest level that I have felt uncertainty about, but no level that I have credence 1 in.) But the *basis* of an agent's felt uncertainty might involve other kinds of uncertainty, and perhaps other epistemic states like ignorance and error as well. So more work would be required to determine what role this kind of uncertainty can play in the subjective levels in the hierarchy.

APPENDIX D.

The Incompleteness Problem and Action-Guidance

The point of a two-level normative theory, it might be thought, is that while the first level tells us what's actually right, the second level is 'action-guiding.' Or: while the first addresses the metaphysical normative questions, the second addresses the epistemological ones. Or: the first level is *theoretical*. It provides necessary and sufficient conditions for right actions. But since we don't always know how to use that theory to derive a prescription for action, we require a *practical* level as well. This level might consist of a set of rules or guidelines that we could follow to ensure that we always act 'as best we can' given our epistemic limitations.

The Incompleteness Argument bears some relation to these claims. But what is it, exactly? I will focus here on the connection to action-guidance. I'll show that the Incompleteness Argument *could* be described as demanding a certain kind of action-guidance from any complete normative theory. (Or from any that can answer subjectivist questions.) However, the kind of action-guidance required is unusual.

But before that, I'll say a bit about the other notions. On my view, subjective normative questions concern the epistemic limitations of normative agents, and answers to them must be sensitive to these limitations. (Or to some of them, at least.) So if this is all that's meant by the metaphysical/epistemological distinction, then I could say that objectivist normative questions are metaphysical while subjectivist ones are epistemological. Likewise, if all that's meant by the theoretical/practical distinction is that practical theories are easier to use to derive prescriptions for action, then I could say that objectivist normative theories are

theoretical while subjectivist ones are practical. And perhaps there's a sense in which, by a 'complete normative theory,' I mean one that would do both the metaphysical/theoretical and the epistemological/practical tasks.

However, I think it's more accurate to present the project I have in mind in a way that could conflict with these distinctions. The kind of normative theories I'm concerned with are theoretical in that they say what's 'actually right,' they cover the normative facts, and they provide necessary and sufficient conditions for right actions (or for what should be done). But on my view there are subjectivist normative facts. There are facts about one should do given one's uncertainty about such-and-such, and these are not identical to facts about what one should do objectively speaking. So one way of thinking about my project is as a theoretical one where the theoretical includes the practical, or at least most of it. (Or as one of answering all the metaphysical normative questions where our epistemic positions create different metaphysical normative questions.) A more accurate thing I can say is that the project involves providing answers to all subjectivist normative questions of the form *what should A do given her uncertainty about p?* where none of these answers must be accessible to A, but where they must still be appropriately sensitive to A's uncertainty. Whether or not these answers could be action-guiding or practical will depend on how much guidance is required by these notions.

As an example of a two-level theory that's designed in order to provide a theoretical level and a practical, or action-guiding, one, consider Fred Feldman's:⁹⁹

Level 1: You morally ought¹ to perform an act if and only if it maximizes utility.

⁹⁹ Feldman 2012 page 166.

Level 2: If you cannot determine what you morally ought¹ to do, then you morally ought to perform an act if and only if it is an outcome of the Utilitarian Decision Procedure.

The Utilitarian Decision Procedure

- Step 1: Consider the acts that you take to be your alternatives—described in “helpful,” “action-guiding” terms;
- Step 2: Consider, insofar as your epistemic state permits, what you take to be their values or perhaps just their relative values;
- Step 3: If you haven’t got useful information about the actual values of your alternatives, then consider how your views about the morality of risk apply to your present situation; and, in light of all this,
- Step 4: Identify the acts in this particular case that seem most nearly consistent with the general policy of maximizing utility where possible while avoiding things that put people at excessive risk of serious harm; and then;
- Step 5: Perform one of them.

Feldman takes pains to make Level 2 a moral theory that offers more guidance than a subjectivist theory like *do what maximizes expected utility*. And the Utilitarian Decision Procedure is definitely helpful. Simply taking the time to complete the first two steps will prove useful in many cases. But unfortunately, it still doesn’t offer guidance in every case.

The main trouble is that the theory is not designed to address truth uncertainty about normative theories.¹⁰⁰ For example, if I’m unsure whether utilitarianism is true, I won’t know whether it makes sense to follow the Utilitarian Decision Procedure or not. I could find myself in a situation where I’ve gone through the steps and know what action I morally

¹⁰⁰ See Smith 2018, Chapter 10 for a thorough criticism of Feldman’s theory that focuses mainly on cases of application uncertainty, as well as for a discussion of the limitations of other theories with a small number of levels.

ought to perform according to Level 2. But the theory still won't answer my question about what I morally ought to do *given my uncertainty about utilitarianism*.

There will also be cases in which one is unable to follow the Utilitarian Decision Procedure. Perhaps I don't know what my views about the morality of risk are or how they apply to my situation. If I can't follow the procedure, then according to the two-level theory there is nothing that I morally ought to do. But this is another reason to doubt the truth of the theory.

So Feldman's two-level theory can't do everything it was designed to do. But what if his decision procedure were expanded significantly? I think a consequence of the Incompleteness Problem is that this would be impossible. However the theory were expanded, we could find a case that would make another addition necessary. If I'm right, then the Incompleteness Problem bears some relation to action-guidance.

A *perfectly* action-guiding normative theory would be such that, no matter what epistemic situation you find yourself in, you can use the theory to determine what you should do. No theory could offer this degree of action-guidance. It's possible, for example, to find yourself in a situation where you are entirely ignorant of the theory, in which case you wouldn't be able to use it. So an 'action-guiding normative theory' usually means something like:

An action-guiding normative theory

A theory that you could always use to determine what to do, *assuming certain background assumptions are met*.

The background conditions include things like: *you know what the theory is and you have the requisite reasoning abilities to use it if supplied with some facts*. They do not usually include things like: *omniscience, certainty about the true moral theory*, etc. Like subjectivity, it makes sense to think

of action-guidance as something that admits of degrees. And, according to the definition just given, the stronger the background conditions would need to be in order for a theory to count as action-guiding, the less action-guidance it provides.

If a theory were perfectly action-guiding, it would answer all subjectivist questions of the form *what should one do, given one's uncertainty about p?* Suppose Nora is uncertain about something. If she's able to use a theory to determine what to do, then the theory's answer to the question *what should Nora do, given her uncertainty?* will be sensitive to what Nora is uncertain about. If the answer were to depend on the truth of anything Nora is uncertain about, then she wouldn't be able to use it to determine what to do.¹⁰¹ So, if a normative theory were perfectly action-guiding, it could be complete.

Would a normative theory that answers all subjectivist questions of the form *what should one do, given one's uncertainty about such-and-such?* be perfectly action-guiding? Not necessarily. If a normative theory can answer all these questions, then it would be sensitive to every possible state of uncertainty. It would be able to give an answer that doesn't rely on what some person is uncertain about, for every possible person and situation. But since there are things besides uncertainty that could prevent someone from being able to use a theory to determine what to do, it wouldn't be perfectly action-guiding. For example, I might have a false belief about something or suffer from an unknown glitch in my reasoning. Even if we

¹⁰¹ This may not be exactly right. In practice, we're less than completely certain about almost everything, and yet we do manage to use some theories to determine what to do. However, that doesn't mean that there aren't many cases in which our uncertainty about something prevents us from being able to determine what to do using some theory. What I propose is that we distinguish between what we're willing to assume for the sake of a decision problem, and what we're not. While we may be uncertain about both kinds of things, it's what we're both uncertain about and unwilling to assume for the sake of a decision problem that prevents us from using theories to determine what to do. For example, in deciding between two restaurants, I may often simplify my decision by assuming that both are open. This doesn't mean that I'm completely certain that they are, but I will act as if I am in making my decision. So when I talk about things that one is uncertain about here, I will mean in particular those things that one is also not willing to assume. This is also the topic of Appendix C.

suppose that the theory can answer all subjectivist questions, it's still not guaranteed to be perfectly action-guiding. Suppose we imagine a very limited agent. Such an agent faces a decision to do A or B, but has no idea what the correct theory is, and has no capacity to understand it or to derive anything from it. The correct theory might still be able to answer the question *what should the very limited agent do, given her extreme limitations?* (Perhaps the answer is that the agent can do either A or B.) But this theory would not be perfectly action-guiding because our extremely limited agent would not be able to use it to determine what she should do.

So a normative theory's ability to answer every subjectivist question of the form *what should one do, given one's uncertainty about such-and-such?* is necessary but not sufficient for it to be perfectly action-guiding. We could, however, think of completeness as requiring a very specific kind of less-than-perfect action-guidance. For example:

A normative theory is **action-guiding*** *iff* if you are uncertain whether p, you could use the theory to determine (with certainty) what you should do given your uncertainty whether p, *assuming that you have no other epistemic limitations besides your state of uncertainty whether p.*

This would be an unusual definition of action-guidingness. Action-guidingness is meant to get at the intuitive idea that a normative theory should be helpful for real agents who are deciding what to do. So action-guiding* will seem too weak. Most real agents will face more trouble than uncertainty whether p in deciding what to do (given their uncertainty whether p). They can be uncertain about other things, they will lack relevant information, and they won't be perfect reasoners. A normative theory that is action-guiding* won't provide any help to these agents.

Imagine that you're in Parfit's Miners Case. You might begin by thinking: *I'm uncertain about where the miners are*. Then you'll be able to use an action-guiding* theory to determine what you should do given your uncertainty about where the miners are. But wait! That's only if you aren't uncertain about anything else. And come to think about it, you may be uncertain about, say, what your evidence is, which kind of moral theory is right, etc. You can add these new things to the statement, but then the theory might rely on something else you're uncertain about in telling you what you should do given your uncertainty about where the miners are, what your evidence is, and which kind of moral theory is right.

In order for a two-level normative theory to avoid the Incompleteness Problem by appealing to action-guidingness, the second level would need to be action-guiding*. But the arguments for Premise 2 of the Incompleteness Argument are also reasons to doubt that any normative theory could be action-guiding*. For example, suppose (for *reductio*) that theory N is action-guiding*. Now imagine that I'm uncertain whether N is true, and that I have no other epistemic limitations. I might be able to use N to figure out what *it* says I should do. But I won't be able to determine with certainty that this is what I should do, because I'd still be uncertain about N.

The requirement that a normative theory be able to answer every question of the form *what should one do, given one's uncertainty about such-and-such?* is best thought of as a weaker and perhaps more fundamental requirement than that a normative theory be action-guiding. It's a necessary condition for a theory's being perfectly action-guiding. But it's also not particularly focused on the issue of whether actual agents would be able to use the theory to determine what to do. Instead, it's focused on a 'theoretical,' as opposed to a 'practical' kind of subjectivist intuition: that there are facts about what agents should do given their

epistemic situations, and that these facts are relevantly sensitive to these epistemic situations whether or not the facts are accessible to the agents themselves.

APPENDIX E.

Extending the *Lite* Hierarchy Theory in Other Directions

The *Lite* Hierarchy Theory could not hope to cover all of normative space, and not just because of the Incompleteness Problem. It doesn't say anything about normative concepts besides SHOULD, and it doesn't say anything about other kinds of subjective normative facts.

In Appendix A, I discuss how the Normativity *Lite* Framework might be extended to cover other 'option-normative' concepts: those that concern or evaluate the options of normative agents. If this can be done, then we could have a different *Lite* Hierarchy Theory for each one of these concepts, and the conjunction of all of these theories could make up the first stage in our open-ended theory.

But it should also be possible to extend the *Lite* Hierarchy Theory to handle cases of ignorance, error, and other epistemic imperfections. The important question is: should answers to questions like *what should one do, given one's **ignorance** about p ?* and *what should one do, given one's **mistaken belief** that p ?* be sensitive to the ignorance and error in the same way that we expect answers to questions like *what should one do, given one's **uncertainty** about p ?* to be sensitive to the uncertainty?

I won't attempt to settle this here. However, if the answer is *yes*, then the extension is mostly straightforward. Ignorance about a normative theory of the form *A should ϕ iff C* can be represented as having no credence or level of confidence in its truth, or as having no credence or level of confidence in whether or not C obtains. Error can be represented as being mistakenly certain of or mistakenly assuming the truth or falsity of the normative theory, or the truth about whether or not C obtains.

Then we'd only need to make a small adjustment to the definitions in the *Lite* Hierarchy Theory:

The Extended *Lite* Hierarchy Theory

Level 1	A OBJECTIVELY SHOULD ϕ = If A were to ϕ , then one of A's most (objectively) valuable reachable states of affairs would obtain.
Level 2	A SUBJECTIVELY ₁ SHOULD ϕ = (i) If A were to ϕ , then one of A's most subjectively ₁ valuable certainly ₁ reachable states of affairs, S, would obtain, and (ii) ϕ is (for A) certainly ₁ linked to S.
...	...

Where:

A's **certainly_n reachable states of affairs**: all states of affairs, S, such that, if A were only uncertain/**ignorant/mistaken** about what she subjectively_{n-1} should do and everything that forms the basis of her uncertainty/**ignorance/error** about what she subjectively_{n-1} should do, a perfect reasoner who's only uncertain/**ignorant/mistaken** about these things (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

ϕ is, for A, **certainly_n linked** to S just in case: if A were only uncertain/**ignorant/mistaken** about what she subjectively_{n-1} should do and everything that forms the basis of her uncertainty/**ignorance/error** about what she subjectively_{n-1} should do, a perfect reasoner who's only uncertain/**ignorant/mistaken** about these

things (and is otherwise fully-informed) could be certain that: if A were to ϕ , then S would obtain.

Subjective_n value: value assigned to states of affairs in a way that is appropriately sensitive to A's uncertainty/**ignorance/error** about what she subjectively_{n-1} should do.

The immediate challenge we'd face if we go this route is that of making sense of cases where an agent is mistaken about what her options are. For example, if I'm mistakenly certain that I can bend space-time and that *if I were to bend space-time thus-and-so, then I'd prevent a dangerous avalanche*, then we might get the result that: I subjectively₁ should bend space-time.

And this would not be a good result. Because I don't have the ability to bend space-time, it's not true that I *should* bend it, in any sense at all. A quick fix is to add in that in cases where the theory would otherwise say that the agent subjectively should perform an impossible action, it says instead that the agent subjectively should *try* to perform that action.

BIBLIOGRAPHY

- Bartha, Paul, John Barker, and Alan Hájek. 2014. Satan, Saint Peter, and Saint Petersburg: Decision Theory and Discontinuity at Infinity. *Synthese* 191 (4): 629-660.
- Baker, Derek. Forthcoming. Skepticism about Ought Simpliciter. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics* 13. Oxford University Press.
- Baker, Derek. 2017. The Varieties of Normativity. In Tristram McPherson, David Plunkett (eds.), *The Routledge Handbook of Metaethics* 567-581.
- Bartha, Paul; Barker, John; and Hájek, Alan. 2014. Satan, Saint Peter, and Saint Petersburg. *Synthese* 191: 629-660.
- Broome, John. 2013. *Rationality through Reasoning*. Wiley-Blackwell.
- Chang, Ruth. 1997. *Incommensurability, Incomparability, and Practical Reason*. Cambridge: Harvard University Press.
- Chang, Ruth. 2007. Incommensurable Values. *Stanford Encyclopedia of Philosophy*. (<https://plato.stanford.edu/entries/value-incommensurable/>)
- Christensen, David. 2013. Epistemic Modesty Defended. In David Christensen and Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*, 77-97.
- Christensen, David. 2010a. Higher-Order Evidence. *Philosophy and Phenomenological Research* 81(1): 185-215.
- Christensen, David. 2010b. Rational Reflection. *Philosophical Perspectives* 24 (1): 121-140.
- Clifford, William K. 1877. The Ethics of Belief. *Contemporary Review* (Jan 1877), collected in Leslie Stephen and Frederick Pollock (eds.), *Lectures and Essays: By the Late William Kingdon Clifford, F.R.S.*, 1886.
- Copp, David. 2009. Toward a Pluralist and Teleological Theory of Normativity. *Philosophical Issues* 19: 21-37.
- Dancy, Jonathan. 2004. *Ethics without Principles*. Oxford: Clarendon Press.
- Dorsey, Dale. 2016. *The Limits of Moral Authority*. Oxford University Press.
- Evers, Daan. 2011. The Standard-Relational Theory of 'Ought' and the Oughtistic Theory of Reasons. *Australasian Journal of Philosophy* 89 (1): 131-147.
- Feldman, Fred. 2012. True and Useful: On the Structure of a Two Level Normative Theory. *Utilitas* 24 2: 151-71.
- Finlay, Stephen. 2014. *Confusion of Tongues: A Theory of Normative Language*. Oxford University Press.
- Finlay, Stephen. 2009. Oughts and Ends. *Philosophical Studies* 143 (3): 315-340.
- Foley, Richard. 1990. Fumerton's Puzzle. *Journal of Philosophical Research* 15: 109-113.
- Fumerton, Richard. 1989. *Reason and Morality*. Cornell University Press.
- Ginet, Carl. 2001. Deciding to Believe. In M. Steup (ed.), *Knowledge, Truth and Duty* 63-76. Oxford University Press.

- Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. (On Formally Undecidable Propositions of Principia Mathematica and Related Systems I.) *Monatshefte für Mathematik*.
- Groenendijk, Jeroen and Stokhof, Martin. 1984. *Studies on the Semantics of Questions and the Pragmatics of Answers*. (Joint Ph.D. Thesis, University of Amsterdam Department of Philosophy.)
- Gustafsson, Johan E., and Olle Torpman. 2014. In Defense of My Favourite Theory. *Pacific Philosophical Quarterly* 95 (2): 159-174.
- Harman, Elizabeth. 2015. The Irrelevance of Moral Uncertainty. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics* 10: 53-79. Oxford University Press.
- Hedden, Brian. 2015. Does MITE Make Right: On Decision-Making Under Normative Uncertainty. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics* 11. Oxford University Press.
- Jackson, Frank. 1991. Decision-Theoretic Consequentialism and the Nearest and Dearest Objection. *Ethics* 101 (3): 461-482.
- Kolodny, Niko and MacFarlane, John. 2010. Ifs and Oughts. *Journal of Philosophy* 107 (3): 115-43.
- Lasonen-Aarnio, Maria. 2015. New Rational Reflection and Internalism about Rationality. *Oxford Studies in Epistemology* 5: 145-179.
- Lasonen-Aarnio, Maria. 2014. Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research* 88 (2): 314-345.
- Lockhart, Ted. 2000. *Moral Uncertainty and Its Consequences*. Oxford University Press.
- MacAskill, William. 2014. *Normative Uncertainty*. (Ph.D. Thesis.)
- Mason, Elinor. 2013. Objectivism and Prospectivism about Rightness. *Journal of Ethics and Social Philosophy* 7 (2): 1-21.
- Nissan-Rozen, Ittay. 2015. Against Moral Hedging. *Economics and Philosophy* 31: 349-369.
- Parfit, Derek. 1988. What We Together Do. (Unpublished.)
- Raatikainen, Panu. 2018. Gödel's Incompleteness Theorems. The Stanford Encyclopedia of Philosophy (Summer 2018 Edition), Edward N. Zalta (ed.), forthcoming URL: <<https://plato.stanford.edu/archives/sum2018/entries/goedel-incompleteness/>>
- Regan, Donald. 1980. *Utilitarianism and Co-operation*. Clarendon Press.
- Rinard, Susanna. 2017. No Exception for Belief. *Philosophy and Phenomenological Research* 94 (1): 121-143.
- Ross, Jacob. 2006a. Rejecting Ethical Deflationism. *Ethics* 116: 742-768.
- Ross, Jacob. 2006b. *Acceptance and Practical Reason*. (Doctoral Dissertation, Rutgers Department of Philosophy.)
- Schroeder, Mark. 2011. Ought, Agents, and Actions. *Philosophical Review* 120 (1): 1-41.
- Schwitzgebel, Eric. 2010. Acting Contrary to our Professed Beliefs, or The Gulf Between Occurrent Judgment and Dispositional Belief. *Pacific Philosophical Quarterly* 91 (4): 531-553.

- Sepielli, Andrew. 2014. Should You Look Before You Leap? *The Philosopher's Magazine* 66: 89-93.
- Sepielli, Andrew. 2013. What to Do When You Don't Know What to Do When You Don't Know What to Do... *Noûs* 47 (1): 521-544.
- Sepielli, Andrew. 2009. What to Do When You Don't Know What to Do. *Oxford Studies in Metaethics* 4: 5-28.
- Smith, Holly. 2018. *Making Morality Work*. Oxford University Press.
- Spencer, Jack and Ian Wells. 2017. Why Take Both Boxes? *Philosophy and Phenomenological Research* 1-22.
- Spencer, Jack and Ian Wells. 2015. The Metaethical Foundations of Decision Theory. (Unpublished.)
- Temkin, Larry. 2012. *Rethinking the Good*. Oxford University Press.
- Tiffany, Evan. 2007. Deflationary Normative Pluralism. *Canadian Journal of Philosophy* 37 (5): 231-262.
- Way, Jonathan and Daniel Whiting. 2016. If You Justifiably Believe That You Ought to Φ , You Ought to Φ . *Philosophical Studies* 173 (7): 1873-1895.
- Weatherson, Brian. 2014. Running Risks Morally. *Philosophical Studies* 167 (1): 141-163.
- Weatherson, Brian. 2010. Do Judgments Screen Evidence? (Unpublished.) (<http://brian.weatherson.org/jse-slides-mit.pdf>)
- Williamson, Timothy. 2008. Why Epistemology Can't be Operationalized. In Q. Smith (ed.) *Epistemology: New Philosophical Essays*. Oxford University Press.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.
- Woods, Jack. Forthcoming. The Authority of Formality. In R. Shafer-Landau (ed.) *Oxford Studies in Metaethics* 13. Oxford University Press.